Dynamics of limit order book: statistical analysis, modelling and prediction

Weibing Huang

THÈSE

présentée pour obtenir

LE GRADE DE DOCTEUR EN SCIENCES DE L'UNIVERSITE PIERRE-ET-MARIE-CURIE

Spécialité : Mathématiques

par

Weibing HUANG

Dynamique des carnets d'ordres: analyse statistique, modélisation et prévision

Soutenue le 15 Dec 2015 devant un jury composé de :
Frédéric Abergel (Ecole Centrale de Paris), Rapporteur
Robert Almgren (Quantitative Brokers, New York University et Carnegie Mellon University), Rapporteur
Aurélien Alfonsi (Ecole des Ponts et Chaussées), Examinateur
Bruno Bouchard (Université Paris Dauphine), Examinateur
Gilles Pagès (Université Pierre et Marie Curie), Examinateur
Mathieu ROSENBAUM (Université Pierre et Marie Curie), Directeur de thèse
Charles-Albert LEHALLE (Capital Fund Management), Directeur de thèse

to Yuan

List of papers being part of this thesis

- HUANG, W., Lehalle, C.-A., and Rosenbaum, M. (2015) *Simulating and analyzing order book data: The queue-reactive model*, *Journal of the American Statistical Association*, 110(509):107-122, 2015.
- HUANG, W. and Rosenbaum, M. (2015) *Ergodicity and diffusivity of Markovian order book models: a general framework*, arXiv preprint, 2015.
- HUANG, W., Lehalle, C.-A., and Rosenbaum, M. (2015) *How to predict the consequences of a tick value change? Evidence from the Tokyo Stock Exchange pilot program*, arXiv preprint, 2015.

Acknowledgement

I am today deeply thankful that I was given the chance of exchanging my views with many intellectual and generous personalities during the past three and a half years. Here I wish to acknowledge those friends and colleagues, without whose support I would never have ended up writing these lines.

In particular, I want to thank the following people:

First and foremost, my advisor Mathieu Rosenbaum, for giving me the chance to pursue a PhD in LPMA. Under his brilliant guidance, I have been given every possible opportunity to develop my research interests. The presence of his ideas and depth of knowledge has inspired many of the works presented in this thesis. This thesis would not have been possible without him.

My advisor at C.A.Cheuvreux, Charles-Albert Lehalle, for introducing me to Mathieu and to the world of research at the University of Pierre and Marie Curie. Charles showed me a first glimpse of the breadth of ideas in the fascinating world of market micro structure and high frequency trading. His enthusiasm for work and capacity of bringing ideas together to solve practical problems showed me an openness to thought that I continue to try to emulate.

Frédéric Abergel and Robert Almgren, for having accepted as referees of this thesis. I am honoured by their lecture of this manuscript and their interest for my work.

Aurélien Alfonsi, Bruno Bouchard and Gilles Pagès for having accepted as examinators of my thesis and for participating the defence of this work.

I have learned much with many friends and colleagues at Cheuvreux and Paris 6. I thank the whole team of C.A. Cheuvreux: Alexandre Denissov, Eduardo Cepeda, Guillaume Pons, Hamza Harti, Joaquin Fernandez-Tapia, Matthieu Lasnier, Minh Dang, Nathanael Mayo, Nicolas Joseph, Paul Besson, Romain Breuil, Silviu Vlasceanu and Stephanie Pelin for their time and disponibility as well as the countless enlightening discussions during these three and a half years work. Special thanks are given to the gaming moments that we shared altogether at the cafeteria and around the babyfoot table. I thank Thibault and Jiatu for the brilliant discussions and for sharing many funny and memoriable moments together.

The secretary of CMAP and LPMA for their disponibility and their helps during these three years.

Last but not least, I want to thank my parents and my wife for their love and unbounded support over the years. I thank Yuan for being on this adventure with me throughout all these years, staying with you always feels like the most beautiful dream I would ever have.

Abstract

This thesis is made of two connected parts, the first one about limit order book modeling and the second one about tick value effects.

In the first part, we present our framework for Markovian order book modeling. The queue-reactive model is first introduced, in which we revise the traditional zero-intelligence approach by adding state dependency in the order arrival processes. An empirical study shows that this model is very realistic and reproduces many interesting microscopic features of the underlying asset such as the distribution of the order book. We also demonstrate that it can be used as an efficient market simulator, allowing for the assessment of complex placement tactics. We then extend the queue-reactive model to a general Markovian framework for order book modeling. Ergodicity conditions are discussed in details in this setting. Under some rather weak assumptions, we prove the convergence of the order book state towards an invariant distribution and that of the rescaled price process to a standard Brownian motion.

In the second part of this thesis, we are interested in studying the role played by the tick value at both microscopic and macroscopic scales. First, an empirical study of the consequences of a tick value change is conducted using data from the 2014 Japanese tick size reduction pilot program. A prediction formula for the effects of a tick value change on the trading costs is derived and successfully tested. Then, an agent-based model is introduced in order to explain the relationships between market volume, price dynamics, bid-ask spread, tick value and the equilibrium order book state. In particular, we show that the bid-ask spread emerges naturally from the fact that orders placed too close to the efficient price have in general negative expected returns. We also find that the bid-ask spread turns out to be the sum of the tick value and the intrinsic bid-ask spread, which corresponds to a hypothetical value of the bid-ask spread under infinitesimal tick value.

Keywords: Limit order book, market microstructure, high frequency data, queuing model, Markov jump process, ergodic properties, volatility, mechanical volatility, market simulator, execution probability, transaction costs analysis, market impact, tick value, market participants' intelligence, priority value, information propagation, equilibrium state.

Contents

Introduction

In this thesis, we aim at building a general mathematical framework for order book modeling which enables us to link the macroscopic features of the price dynamics with the microscopic properties of the underlying asset. On the one hand, we want to shed light on some of the fundamental issues in order book modeling, such as the ergodicity of the order book and the role played by the tick value. On the other hand, our goal is also to provide relevant tools for market participants and regulators, helping them analyzing complex trading algorithms or effects of some regulatory measures.

Motivations

Can we explain the price dynamics of a stock from a microstructural point of view? This question is at the heart of market microstructure literature for decades. Many interesting results have been obtained and are nowadays used by practitioners as theoretical guidelines in the three main branches of high frequency trading: optimal execution, market making and statistical arbitrage. Such results usually aim at building a link between the high frequency dynamics of the the asset and the well established macroscopic features of the underlying stock. One of the most fascinating challenges in this context is to explain those links starting from the very finest microstructural scale, that of the limit order book. This limit order book modeling problem consists in establishing a tractable and relevant mathematical formulation for the market mechanics of order matching and order queueing, and for the complex behavior of market participants.

Although important advances have been made in the recent years about limit order book modeling, market participants still often find themselves in lack of applicable models when dealing with practical problems. In particular, in strong contrast to empirical findings, most existing limit order book models use an homogenous Poisson assumption on the order arrival process. Moreover, several models consider only the dynamics at the best bid/ask limits or assume constant bid-ask spread. These simplifications largely reduce the applicability of such models, as many algorithms used in practice operate on non-best limits and are very sensitive to their orders' priority in the queue. Our goal being to provide relevant tools for market participants and regulators, the first question we want to address in this thesis is obviously the following one:

Question 1. *How to build realistic order book models?*

At the microscopic level, market activities are random and unpredictable. Yet when properly scaled in time, they exhibit many interesting regularities. For example, the price dynamics at the macroscopic level can be quite well approximated by a Brownian motion and the limit order book's average shape tends to follow the same distribution across different trading days.

This last regularity is closely related to the ergodicity of the limit order book system. In our first chapter, we introduce a state-dependent order book model, the queue-reactive model, that is particularly suitable for large tick assets[1]. This model is then extended to a more general Markovian framework, enabling us to deal with small tick assets and including most of the classical order book models. Thus we want to give an answer to the following important theoretical question for Markovian order book modeling:

Question 2. *In a general Markovian framework for limit order book modeling, what are the required conditions to obtain ergodic dynamics?*

In our theoretical Markovian framework, market participants react differently under different market conditions. In the empirical study associated to the queue-reactive model, we find a strong contrast in traders behavior given various order book states, which validates our intuition that this state-dependent hypothesis is far more realistic than the traditional Poisson assumption. Interestingly, the estimated intensity functions for various assets for limit/market order insertion and limit order cancellation share many similarities. For example, the cancellation intensities are all found to be concave functions of the queue size, and the limit order insertion intensities at best limits tend to be essentially increasing functions. These common patterns can be seen as results of market participants intelligence. While they may be explained qualitatively by intuitive arguments such as the existence of market priority and risk of overrun, the following question remains very intricate:

Question 3. *How to get a quantitative agent-based approach enabling us to understand the behavior of market participants towards various states of the book and to retrieve the most important limit order book features?*

One microstructural parameter having a strong influence on the trading practice of market participants is the tick value. We are particularly interested in this quantity since it is often considered the most relevant device to regulate the behavior of high frequency traders and to control market efficiency. Although several tick value change programs have been conducted in the recent years, most of them are designed using only empirical analysis and focus on the outcomes of the tick value modification in an *ex post* basis. Hence the effects of tick value changes have not been really understood and prediction tools enabling us to forecast the consequences of a tick value change are missing. Here we wish to fill this gap, with the aim to help market regulators to better determine the target tick values of such programs. Therefore we are considering the following question:

Question 4. *How to predict the consequences of tick value changes?*

Outline

This thesis is made of two main parts: order book modeling and tick size effects. Each question presented above corresponds to a chapter in one of these two parts.

In Part I, we present our work on order book modeling. In Chapter I, we answer Question 1 by building the queue-reactive model, in which state dependency is included in the order flow dynamics. We propose to split the order book modeling issue into two steps: i) dynamics of the order arrival processes around a constant reference price; ii) dynamics of the reference price. This enables us to design three different versions of the model, with various hypotheses on the

[1]A large tick asset is defined as an asset whose bid-ask spread is almost always equal to one tick.

information set used by market participants when making their trading decisions. Unlike the traditional Poisson assumption, order arrival intensities are assumed to be functions of the state of the order book. Estimation methods are provided and empirical studies are conducted on some large tick assets, showing that many macroscopic features can be explained adding state dependency in the order book dynamics. In particular, the distribution of the order book's state is very well explained in this framework. We then show how to use this model as a market simulator for analyzing complex trading algorithms.

The answer to Question 2 is given in Chapter II. We first extend the queue-reactive model to a general Markovian framework, including most classical order book models. This new framework, while respecting the double-auction mechanism of the order book, imposes very little constraints. For example, the order size is allowed to be random, with distribution depending again on the order book's state, and the reference price dynamics can also be linked to order book information such as the bid-ask imbalance. Ergodicity conditions in this framework are then discussed in details. Essentially, we show that if the incoming order flow (that is the limit order insertion rate) does not exceed the outgoing order flow (that is the sum of the market order and cancellation rates), then the order book state is an ergodic process. Furthermore, the price dynamics converges to a Brownian motion when properly rescaled.

Results about the role of the tick value are presented in Part II. To answer Question 4, we conduct in Chapter III an empirical study of the effects of tick value changes based on data from the Japanese tick size reduction pilot program between June 2013 and July 2014. We demonstrate that the approach introduced in Dayri and Rosenbaum (2012) allows for an *ex ante* assessment of the consequences of a tick value change on the microstructure of an asset. We focus on forecasting the future costs of market and limit orders after a tick value change and show that our predictions are very accurate. Furthermore, for each asset involved in the pilot program, we are able to define *ex ante* an optimal tick value.

We finally present in Chapter IV a preliminary attempt to answer Question 3. We split market participants into three types: informed traders, noise traders and market makers. Informed traders receive market information such as the current efficient price, which is hidden to noise traders. Market makers have also access to this information but with some delay, and they place limit orders when it is profitable. In a first model, we consider an idealized setting where the tick value constraint is removed, and assume that both informed traders and market makers have an infinite reaction speed to new information. In that case, we obtain a link between the price dynamics, the market volume and the equilibrium order book shape. We then study the effects of introducing the tick value constraint. We find that when the traded price becomes discrete, the priority value of a limit order can be properly defined and computed. Furthermore, the consequences of the uncertainty faced by market makers about the efficient price are discussed in our framework. We also provide insights on how a new piece of information is digested and propagated between informed traders and market makers and on the speed of order book recovery after a transaction.

We now give a rapid overview of the main results obtained in this thesis.

1 Part I: Limit Order Book Modeling

Understanding the limit order book dynamics is one of the fundamental issues in modern electronic financial markets. Many practical problems, such as the design of a realistic market simulator and the performance evaluation of a high frequency trading algorithm, rely heavily on a reasonable limit order book model. Existing models often assume zero intelligence for market participants and focus only on dynamics at best limits, see for example Smith, Farmer, Gillemot, and Krishnamurthy (2003) and Cont and De Larrard (2013). This largely reduces their appeal for practice. In this part, we aim at building a complete order book model in which limits several ticks away from the best ones are considered and where market participants act in an intelligent way towards various order book states.

In Chapter I, we introduce the queue-reactive model for order book dynamics. The key idea in this model is to split the order book modeling issue into two parts: the order arrival dynamics during period of constant reference price and the dynamics of the reference price. This approach enables us to deal with the strong dependencies between the different limits and to study the sensitivity of the trading activities towards various order book states. We show that the queue-reactive model explains very well the asymptotic distribution of the order book and demonstrate its applicability to assess complex trading algorithms by conducting a detailed analysis of two order placement tactics. The ergodicity conditions of an extended Markovian order book framework are discussed in Chapter II. We prove that the rescaled price process converges to a Brownian motion and the order book state to an invariant distribution under some very general assumptions.

1.1 The Queue-reactive Model

1.1.1 Dynamics of the limit order book in a period of constant reference price

We model the limit order book as a $2K$ dimensional vector, where K denotes the number of available limits on the bid and ask side. By defining the reference price p_{ref} as the center of these $2K$ limits and assuming it is constant, the limit order book dynamics can be described by a continuous time Markov jump process $X(t) = (Q_{-K}(t), ..., Q_{-1}(t), ..., Q_1(t), ..., Q_K(t))$, where $Q_i(t)$ is the number of available orders at the i-th limit. The quantity p_{ref} can be viewed as some current consensus price level and is used to index the limits. Three types of orders are considered: limit orders, cancellations and market orders, and their sizes are assumed to be constant for each limit (we set them here to one for simplicity). Under these assumptions, the infinitesimal generator matrix $\mathcal{Q}_{x,y}$ of the process $X(t)$ can be written as follows ($e_i = (a_{-K}, ..., a_i, ..., a_K)$, where $a_j = 0$ for $j \neq i$ and $a_i = 1$):

$$\begin{aligned}
\mathcal{Q}_{q,q+e_i} &= f_i(q) \\
\mathcal{Q}_{q,q-e_i} &= g_i(q) \\
\mathcal{Q}_{q,q} &= -\sum_{p\in\Omega, p\neq q} \mathcal{Q}_{q,p} \\
\mathcal{Q}_{q,p} &= 0, \text{otherwise.}
\end{aligned}$$

Ergodicity conditions

Write Ω for the state space of q. The two following assumptions are needed for the ergodicity of the process $X(t)$:

Assumption 1. *(Negative individual drift) There exist a positive integer C_{bound} and $\delta > 0$, such that for all i and all $q \in \Omega$, if $q_i > C_{bound}$,*

$$f_i(q) - g_i(q) < -\delta.$$

Assumption 2. *(Bound on the incoming flow) There exists a positive number H such that for any $q \in \Omega$,*

$$\sum_{i \in [-K,\dots,-1,1,\dots,K]} f_i(q) \leq H.$$

The first assumption states that the queue size of a limit should have a tendency to decrease when it becomes too large, while the second one ensures no explosion in the system. Under these two assumptions, we have the following ergodicity result for the $2K$-dimensional queuing system with constant reference price, which will be the basis for our asymptotic study as well as for the estimation procedures.

Theorem 1. *Under the above two assumptions, the $2K$-dimensional Markov jump process $X(t)$ is ergodic.*

Ergodicity conditions are discussed in more details in Chapter II.

The functions f_i and g_i model the state dependency of market participants' behavior. Then different assumptions on the information set used by traders lead to different models in the above framework. Three models are proposed to describe the order book dynamics under constant reference price.

Model I: Collection of independent queues

In Model I, we assume independence between the flows arriving at different limits:

$$\begin{aligned} f_i(q) &= \lambda_i^L(q_i) \\ g_i(q) &= \lambda_i^C(q_i) + \lambda_i^M(q_i), \end{aligned}$$

where λ_i^L, λ_i^C, λ_i^M correspond to the intensities of limit orders, cancellations and market orders respectively.

Model I enables us to study the influence of the target queue size on market participants' behavior. In our empirical study conducted on large tick stocks, we find the following interesting repetitive patterns on the intensity functions (we take $K = 3$ in our experiments, the estimated intensities for the stock France Telecom are shown in Figure .1):

- Limit order insertion:

  $Q_{\pm 1}$: The intensity of the limit order insertion process is approximately a constant function of the queue size, with a significantly smaller value at 0. This can be explained by the fact that creating a new best limit is viewed as risky (inserting a limit order in an empty queue creates a new best limit and the market participant placing this order is the only one standing at this price level).

  $Q_{\pm 2,3}$: The intensity is approximately a decreasing function of the queue size. This interesting result probably reveals a quite common strategy used in practice: posting orders at the second limit when the corresponding queue size is small to seize priority.

- Limit order cancellation:

  $Q_{\pm 1}$: In contrast to the classical hypothesis of linearly increasing cancellation rate, see for example Cont, Stoikov, and Talreja (2010), the intensity of order cancellation is found to be an increasing concave function for $Q_{\pm 1}$. Such result can be explained by the

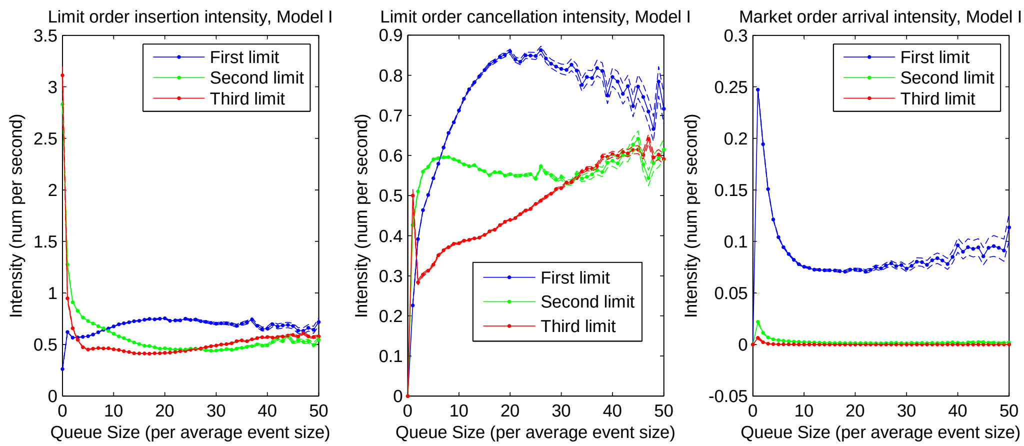

Figure .1: Intensities at $Q_{\pm i}$, $i = 1,\ 2,\ 3$, France Telecom

existence of the priority value, that is the advantage of a limit order compared with another limit order standing at the rear of the same queue. Actually, orders with lower priority are more likely to be canceled, see Gareche, Disdier, Kockelkoren, and Bouchaud (2013).

$Q_{\pm 2}$: The rate of order cancellation attains more rapidly its asymptotic value, which is lower than that for $Q_{\pm 1}$. Compared to the first limit case, market participants at the second limit have even stronger intention not to cancel their orders when the queue size increases. This is probably due to the fact that these orders are less exposed to short term market trends than those posted at $Q_{\pm 1}$ (since they are covered by the volume standing at $Q_{\pm 1}$ and their price level is farther away from the reference price).

$Q_{\pm 3}$: The priority value is smaller at the third limit since it takes longer time for $Q_{\pm 3}$ to become the best quote if it does. The rate of order cancellation increases almost linearly.

- Market orders:

$Q_{\pm 1,2,3}$: The rate decreases exponentially with the available volume at $Q_{\pm 1,2,3}$. This phenomena is easily explained by market participants "rushing for liquidity" when liquidity is rare, and "waiting for better price" when liquidity is abundant.

In Model I, each queue is actually a birth-death process whose invariant distribution can be computed explicitly: denote by π_i the stationary distribution of the limit Q_i, and define the arrival/departure ratio vector ρ_i by

$$\rho_i(n) = \frac{\lambda_i^L(n)}{(\lambda_i^C(n+1) + \lambda_i^M(n+1))}.$$

Then we have:

$$\begin{aligned} \pi_i(n) &= \pi_i(0) \prod_{j=1}^{n} \rho_i(j-1) \\ \pi_i(0) &= \Big(1 + \sum_{n=1}^{\infty} \prod_{j=1}^{n} \rho_i(j-1)\Big)^{-1}. \end{aligned}$$

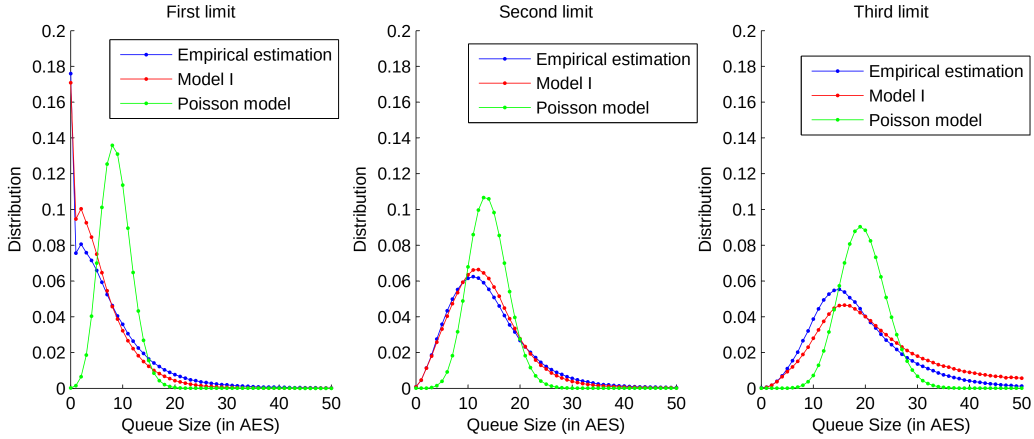

Figure .2: Model I, invariant distributions of $q_{\pm1}, q_{\pm2}, q_{\pm3}$, France Telecom

In Figure .2, we compare the theoretical asymptotic distributions with the empirical distributions observed at $Q_{\pm1}, Q_{\pm2}, Q_{\pm3}$, and with the invariant distributions from a Poisson model with constant limit/market order arrival rate and linear cancellation rate. The theoretical asymptotic distributions are found to be very good approximations of the empirical ones estimated from market data. This suggests that the empirical order book shape can be explained by the asymptotic equilibrium of order flow dynamics with state dependency.

Model II: Dependent case

In the dependent case, we differentiate best and non-best limits and also add dependence between the bid and ask limits. The generator of the process takes the following form:

$$\begin{aligned} f_i(q) &= \lambda_i^L(q) \\ g_i(q) &= \lambda_i^C(q) + \lambda_{buy}^M(q)\mathbf{1}_{bestask(q)=i}, \text{if } i > 0 \\ g_i(q) &= \lambda_i^C(q) + \lambda_{sell}^M(q)\mathbf{1}_{bestbid(q)=i}, \text{if } i < 0. \end{aligned}$$

Model II^a: Two sets of dependent queues

In Model II^a, we propose to consider $\lambda_{\pm2}^L$ and $\lambda_{\pm2}^C$ as functions of $q_{\pm2}$ and $\mathbf{1}_{q_{\pm1}>0}$. Intensities at $Q_i, i \neq \pm2$ remain functions of q_i only. Thus, in the empirical study of Model II^a, we focus on understanding market participants' behavior under two different situations: $q_{\pm1} = 0$ and $q_{\pm1} > 0$. One of our interesting findings is the following one on the limit order arrival process:

- Limit order insertion: both intensities are decreasing functions of the queue size. In the first case ($q_{\pm1} = 0$), the limit order insertion intensity reaches very rapidly its asymptotic value. In the second case ($q_{\pm1} > 0$), the intensity starts at a higher value for $q_2 = 0$ but continues to go down to a much lower value. This is likely related to the following arbitrage strategy: post passive orders at a non-best limit when its size is small, wait for this limit to eventually become the best limit and then gain the profit from having the priority value. For example, when the considered limit becomes the best one, one can decide to stay in the queue if its size is large enough to cover the risk of short term market trend, or to cancel the orders if the queue size is too small.

The joint asymptotic distribution for the limit order book state (q_1, q_2) can be computed numerically in Model II^a, using the fact that it is a quasi-birth-and-death process, see Latouche and Ramaswami (1999), and is found to be again a very good approximation of the empirical one.

Model II^b: Modeling the bid-ask dependences

To study the interactions between the bid side and the ask side, we define the function $\mathscr{S}_{m,l}(x)$ for representing four different ranges of values for the queue sizes (empty, small, usual and large):

$$\begin{aligned}
\mathscr{S}_{m,l}(x) &= \mathscr{Q}^0 \text{ if } x = 0 \\
\mathscr{S}_{m,l}(x) &= \mathscr{Q}^- \text{ if } 0 < x \leq m \\
\mathscr{S}_{m,l}(x) &= \bar{\mathscr{Q}} \text{ if } m < x \leq l \\
\mathscr{S}_{m,l}(x) &= \mathscr{Q}^+ \text{ if } x > l,
\end{aligned}$$

for well chosen m and l.

In Model II^b, market participants at $Q_{\pm 1}$ adjust their behavior not only according to the target queue size, but also to the size of the opposite queue. The rates $\lambda^L_{\pm 1}$ and $\lambda^C_{\pm 1}$ are thus modeled as functions of $q_{\pm 1}$ and $\mathscr{S}_{m,l}(q_{\mp 1})$. Regime switching at $Q_{\pm 2}$ is kept in this model: $\lambda^L_{\pm 2}$, $\lambda^C_{\pm 2}$ are assumed to be functions of $q_{\pm 2}$ and $\mathbf{1}_{q_{\pm 1} > 0}$.

Some remarks are in order:

- Limit order insertion: the limit order insertion rate is a decreasing function of the opposite queue size. In particular, when the opposite queue is empty, it is significantly larger. Indeed, in that case, the “efficient" price is likely to be closer to the opposite side. Therefore limit orders at the non empty first limit are likely to be profitable.

- Limit order cancellation: the cancellation rates for different ranges of Q_{-1} are similar in their forms but have different asymptotic values. This rate is not surprisingly a decreasing function of the liquidity level on the opposite side. Indeed, when this level becomes low, many market participants cancel their limit orders and send market orders since the market is likely to move in an unfavorable direction.

- Market orders: we see that when the liquidity available on the opposite side is abundant, more market orders are sent. Indeed, in that case, transactions at the target queue are relatively cheap as its price level is temporarily closer to the efficient price. In the special situation $q_{-1} = 0$, the price level at Q_1 can seem relatively attractive since it is much closer to the reference price than the opposite best price, which is in that case 1.5 ticks away from it. This explains why the market order intensity is larger when the opposite queue is empty than when its size is small.

Invariant distributions for the limit order book in Model II^b can be obtained using Monte-Carlo simulations, and the comparison results with the empirically estimated ones are still very satisfactory.

1.1.2 A time consistent model with stochastic limit order book and dynamic reference price

We then propose a model accommodating a dynamic limit order book center: the queue-reactive model.

The queue-reactive model

Dynamics of the reference price is added by linking p_{ref} with the mid price p_{mid}: we assume that changes of p_{ref} are triggered by changes of the mid price with some probability θ. We also add another parameter θ^{reinit} to incorporate price jumps resulting from external information: in such case, the order book state is redrawn from some invariant distribution around the new reference price. These two parameters are calibrated using 10 min price volatility and the mean reversion ratio η from Robert and Rosenbaum (2011).

Maximum mechanical volatility

When $\theta^{reinit} = 0$, price fluctuations are only endogenously generated by the order book dynamics. In such case, the volatility is an increasing function of θ and attains its maximum value when $\theta = 1$. The associated volatility is called maximum mechanical volatility, and is found to be often smaller than the empirical volatility, which justifies the use of the parameter θ^{reinit}.

1.1.3 Order placement analysis

In practice, an execution algorithm gives answers to the two following questions:

- Order Scheduling: how to distribute the target volume across the trading horizon?
- Order Placement: how to send individual orders to the order book?

The first question is widely studied in the literature, see for example Bertsimas, Lo, and Hummel (1999); Almgren and Chriss (2000); Bouchard, Dang, and Lehalle (2011). Answers to it often rely on some "optimal trading curve", which depends mainly on intraday factors such as the average volume curve, the intraday volatility and the average market impact profile. The second question can be seen as the microstructural version of the first one, but is much more difficult to solve since the dynamics are more complex. Related academic works address the problem of determining the optimal order type (whether to send limit or market order) Harris and Hasbrouck (1996), or of finding the best position to place the order Laruelle, Lehalle, and Pagès (2013). In practice, order placement tactics are usually much more complex.

While most existing approaches for post-trade performance analysis focus on the overall performance, it is actually more reasonable to separate the order scheduling part from the order placement part. Performance of order placement tactics depends more on ultra-high frequency features such as the latency, the queue priority, bid-ask imbalance, etc, which have generally little influences over the choice of the "optimal trading curve". Moreover, the same order scheduling strategy can be coupled with different order placement tactics to build different execution algorithms. In such cases, it is important to be able to understand the pros and cons of each placement tactic so that an informed choice can be made to determine the best tactic under different market conditions.

We present in this introduction an application example to show how the queue-reactive model can be used in the context of order placement analysis for sophisticated tactics.

Denote n_{total} for the total quantity to execute and M for the number of trading slices. An order scheduling strategy gives the target quantity to be executed in each slice, denoted by n_i ($\sum_{i=1}^{M} n_i = n_{total}$). Two types of order scheduling strategies, denoted by **S**1 and **S**2, are considered in this example:

- **S**1: A linear scheduling ($n_i = n_{total}/M$), used for the VWAP benchmark (volume weighted average price).
- **S**2: An exponential scheduling $n_i = n_{total}(e^{-(i-1)/4} - e^{-i/4})$, used for the benchmark S_0 (arrival price).

An order placement tactic can be seen as a predefined procedure of order management, ensuring the execution of the target quantity within the slice. The following two tactics will be considered in our analysis: in the i-th slice, both tactics post a limit order of size n_i at the best offer queue at the beginning of the period, and send a market order with all the remaining quantity to complete the execution of the target volume at the end time of the slice. In between:

- **T**1 (Fire and forget): When p_{mid} (the mid price) changes, cancel the limit order and send a market order on the opposite side with all the remaining volume if any.
- **T**2 (Pegging to the best): When the best offer price changes or our order is the only remaining order at the best offer limit, cancel the order and repost all the remaining volume at the newly revealed best offer queue.

Performance measure

To understand the effects of order placement tactic on the execution's slippage, we propose the two following measures on an execution's performance: Slippage and Slippage^{theo}.

$$\begin{aligned} \text{Slippage} &= \frac{P_{benchmark} - P_{exec}}{P_{benchmark}}, \\ \text{Slippage}^{theo} &= \frac{P_{benchmark} - P_{exec}^{theo}}{P_{benchmark}}, \end{aligned}$$

where $P_{exec}^{theo} = \sum_{i=1}^{M} n_i \text{VWAP}^i$ represents the average execution price if the algorithm obtains the same price as the market VWAP in each trading slice. Essentially, Slippage measures the overall performance of the execution algorithm as a combination of order placement tactic and order scheduling strategy, while Slippage^{theo} measures the quality of the scheduling strategy alone and neglects the randomness in executed price due to the order placement tactic in each trading slice.

2000 simulations are launched for each couple of (**S**1/**S**2, **T**1/**T**2). We then estimate the probability density functions of Slippage^{theo} and Slippage. The results are shown in Figure .3. The simulation results suggest that the same order scheduling strategy can have very different performance when being coupled with different order placement tactics: **T**2 ("Pegging to the best") performs better than **T**1 ("Fire and forget") when being coupled with a linear scheduling strategy with VWAP benchmark, while **T**1 slightly outperforms **T**2 when an exponential scheduling strategy with arrival price benchmark is considered. By executing most of the target quantity via limit orders, **T**2 obtains on average a better price than that of a more market orders based

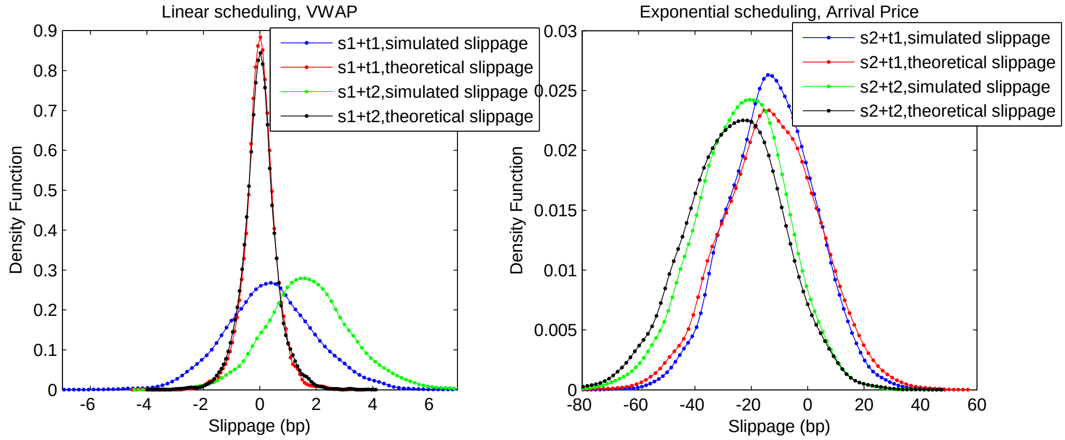

Figure .3: Simulation results for the tactics

tactic. However, at the same time, it creates a larger impact than **T1** since the order stays longer in the queues. Note that market impact profiles for these two tactics can also be obtained using Monte-Carlo simulations.

1.2 A General Framework for Markovian Order Book Modeling

In Chapter II, we extend the queue-reactive model to a general Markovian framework.

1.2.1 Order book dynamics

We represent the order book $X(t)$ by two elements: its center position denoted by p_{ref} (which plays the same role as p_{ref} in the queue-reactive model) and its form $[q_{-K}, ..., q_{-1}, q_1, ..., q_K]$. The use of one unique reference price that is not directly observable from the order book state gives us flexibility for modeling the order book and enables us to differentiate two types of jumps in the order book dynamics: pure order book state jumps (for which the order book center p_{ref} stays invariant) and common jumps (jumps in which a reference price change is involved).

Pure order book jump

We assume that a pure order book jump can only happen at one specific queue at each jump time. Unlike the queue-reactive model, buy/sell limit orders are allowed to be inserted on both parts of the reference price. Moreover, the jump size is now random. Thus, in term of generator, we have, with $2K$ functions f_i, g_i:

$$\begin{aligned}
\mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q,n) \\
\mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q,n) \\
\tilde{\mathcal{Q}}_{(q,p),(q',p)} &= 0, \text{ otherwise.}
\end{aligned}$$

Common jump

New information such as the arrival of a market order may affect the value of the consensus price, and such effect takes place with some delay in practice. In our framework, we use a discretized p_{ref} (with tick value denoted by α) and model the jump rate of p_{ref} as function of the order book state $q(t)$:

$$\begin{aligned}
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p+\alpha)} &= u(q) \\
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p-\alpha)} &= d(q) \\
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p\pm n\alpha)} &= 0, \text{ for } n \geq 2.
\end{aligned}$$

Note that when the order book center changes, the values of q_i switches immediately to the value of one of its neighbors. We thus introduce two boundary distributions π_{-K} and π_K for generating new queue sizes at $Q_{\pm K}$ as we keep only K limits on each side. As in the queue-reactive model, we assume that $q_{\pm K}$ is redrawn from some distribution (π^{inc} if p_{ref} increases, π^{dec} if it decreases) with some probability θ^{reinit} whenever a reference price jump happens. For $q \in \Omega$ (Ω denotes the state space of all possible order book shape) and $l \in \mathbb{R}$, write $q^+ = [q_{-K},...,q_{-1},q_1,...,q_{K-1}]$, $q^- = [q_{-K-1},...,q_{-1},q_1,...,q_K]$, $[q^+,l] = [q_{-K},...,q_{-1},q_1,...,q_{K-1},l]$ and $[l,q^-] = [l,q_{-K-1},...,q_{-1},q_1,...,q_K]$. We have for any q,q',q'' such that $q'^+ \neq q^+$ and $q''^- \neq q^-$:

$$\begin{aligned}
\mathcal{Q}_{(q,p),([q^+,l],p+\alpha)} &= (1-\theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+,l]) \\
\mathcal{Q}_{(q,p),(q',p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\
\mathcal{Q}_{(q,p),([l,q^-],p-\alpha)} &= (1-\theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l,q^-]) \\
\mathcal{Q}_{(q,p),(q'',p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q'').
\end{aligned}$$

The infinitesimal generator matrix of the order book process

Gathering all the above hypotheses together, we obtain the following description for the infinitesimal generator matrix of the Markovian jump process $X(t)$:

Assumption 3. *For any $q,q',q'',\tilde{q} \in \Omega$, $p,\tilde{p} \in \alpha(0.5+\mathbb{Z})$, $n \in \mathbb{N}^+$, $l \in \mathbb{Z}$, such that $q'^+ \neq q^+$ and $q''^- \neq q^-$, the infinitesimal generator matrix $\mathcal{Q}$ of the process $X(t)$ is of the following form (with $2K$ functions $f_i, g_i : \Omega \times \mathbb{N}^+ \to \mathbb{R}^+$ and 2 functions $u,d : \Omega \to \mathbb{R}^+$) :*

$$\begin{aligned}
\mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q,n) \\
\mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q,n) \\
\mathcal{Q}_{(q,p),([q^+,l],p+\alpha)} &= (1-\theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+,l]) \\
\mathcal{Q}_{(q,p),(q',p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\
\mathcal{Q}_{(q,p),([l,q^-],p-\alpha)} &= (1-\theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l,q^-]) \\
\mathcal{Q}_{(q,p),(q'',p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q'') \\
\mathcal{Q}_{(q,p),(q,p)} &= -\sum_{(\tilde{q},\tilde{p})\in\Omega\times\alpha(0.5+\mathbb{Z}),(\tilde{q},\tilde{p})\neq(q,p)} \mathcal{Q}_{(q,p),(\tilde{q},\tilde{p})} \\
\mathcal{Q}_{(q,p),(\tilde{q},\tilde{p})} &= 0, \textit{otherwise}.
\end{aligned}$$

Note that up to minor modifications, most classical order book models such as the zero-intelligence model of Smith et al. (2003) and that of Cont et al. (2010), and the queue-reactive model, can be included in this framework.

1.2.2 V-Uniform ergodicity

Constant reference price

We now discuss ergodicity conditions for the Markovian process $X(t)$. V-uniform ergodicity implies the existence of a unique invariant distribution for the state vector $q(t)$, which is very useful in explaining the empirical order book distribution, as we have already seen in the queue-reactive model. We write

$$f_i^*(q) := \sum_n f_i(q,n),$$

$$g_i^*(q) := \sum_n g_i(q,n),$$

and consider two probability measures on $\mathbb{N}^+$

$$l_i(q,n) := \frac{f_i(q,n)}{f_i^*(q)},$$

$$k_i(q,n) := \frac{g_i(q,n)}{g_i^*(q)},$$

and their related moment-generating functions $G^{f,i,q}(z)$ and $G^{g,i,q}(z)$. We make the following assumptions:

Assumption 4. *For any order book state q and any $i \geq i_{bestask}$, $g_i(q,n) = 0$ for any $n > q_i$ and for any order book state q and any $i \leq i_{bestbid}$, $f_i(q,n) = 0$ for any $n > -q_i$.*

Assumption 5. *There exists $z^* > 1$ such that for any q and i, $G^{f,i,q}(z^*) < \infty$ and $G^{g,i,q}(z^*) < \infty$. Furthermore, there exists $L > 0$ such that for any i,*

$$\overline{\lim_{z\to 1^+}} \sup_q [f_i^*(q)G^{f,i,q}(z)\mathbf{1}_{i>i_{bestbid}} + g_i^*(q)G^{g,i,q}(z)\mathbf{1}_{i<i_{bestask}}] < L.$$

Assumption 6. *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned}
\overline{\lim_{z\to 1^+}} \sup_{(q,i):q_i>U, i\geq i_{bestask}} [f_i^*(q) - g_i^*(q)\frac{1-G^{g,i,q}(z^{-1})}{G^{f,i,q}(z)-1}] &< -r \\
\overline{\lim_{z\to 1^+}} \sup_{(q,i):q_i<-U, i\leq i_{bestbid}} [g_i^*(q) - f_i^*(q)\frac{1-G^{f,i,q}(z^{-1})}{G^{g,i,q}(z)-1}] &< -r.
\end{aligned}$$

Assumption 7. *For any $z > 1$,*

$$\begin{aligned}
B_f(z) &:= \inf_{(q,i):q_i>U, i\geq i_{bestask}} (G^{f,i,q}(z)-1) > 0 \\
B_g(z) &:= \inf_{(q,i):q_i<-U, i\leq i_{bestbid}} (G^{g,i,q}(z)-1) > 0.
\end{aligned}$$

Under these assumptions, we have the following result:

Theorem 2. *When $u = d \equiv 0$, under Assumption 4, 5, 6 and 7, the continuous-time Markov jump process $q(t)$ is non-explosive, V-uniformly ergodic and positive Harris recurrent.*

For the embedded Markov chain $q(n)$, defined as $q(n) = q(J_n)$, with J_n the time of the n-th jump, and $q(J_n)$ the state of the LOB after this event, a different assumption is needed for its ergodicity. We write

$$a_i^*(q) = \frac{f_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]}, \; b_i^*(q) = \frac{g_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]},$$

for the proportions of queue size increases and decreases, and replace Assumption 6 by the following one.

Assumption 8. *There exist* $r > 0$ *and* $U > 1$ *such that*

$$\begin{aligned}
\overline{\lim_{z \to 1^+}} \sup_{(q,i):q_i > U, i \geq i_{bestask}} [a_i^*(q) - b_i^*(q) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1}] &< -r \\
\overline{\lim_{z \to 1^+}} \sup_{(q,i):q_i < -U, i \leq i_{bestbid}} [b_i^*(q) - a_i^*(q) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1}] &< -r.
\end{aligned}$$

Then we can prove the following theorem:

Theorem 3. *When* $u = d \equiv 0$, *under Assumptions 4, 5, 7 and 8, the embedded discrete-time Markov chain* $q(n)$ *is V-uniformly ergodic and positive Harris recurrent.*

The main idea to prove the ergodicity of q is to design an appropriate Lyapunov function V, on which the following negative drift condition is satisfied for some $\gamma > 0$ and $B > 0$:

$$\begin{aligned}
\mathcal{Q}V(q) &:= \sum_{q' \neq q} \mathcal{Q}_{qq'}[V(q') - V(q)] \\
&\leq -\gamma V(q) + B.
\end{aligned}$$

Then the above theorems can be derived using Theorem 4.2 and Theorem 6.1 in Meyn and Tweedie (1993). Note that the same kind of method is used in Abergel and Jedidi (2011) in order to show ergodicity properties of zero-intelligence order book models.

General case

For some $z, U > 1$, let

$$V^z([q,c]) = \sum_{i=-K, i \neq 0}^{K} z^{|q_i| - U}.$$

When p_{ref} is no longer constant, we make the two following additional assumptions:

Assumption 9. *There exist* $z > 1$ *and* $L^\pi > 0$ *such that for* Q^{inc}, Q^{dec}, Q_K, Q_{-K} *four random variables such that* $Q^{inc} \sim \pi^{inc}$, $Q^{dec} \sim \pi^{dec}$, $Q_K \sim \pi_K$ *and* $Q_{-K} \sim \pi_{-K}$:

$$\mathbb{E}[V^z([Q^{inc}, c])] + \mathbb{E}[V^z([Q^{dec}, c])] + \mathbb{E}[z^{|Q_K| - U}] + \mathbb{E}[z^{|Q_{-K}| - U}] \leq L^\pi.$$

Assumption 10. *There exists a finite set* $W \subset \Omega$ *such that the upper bound of the proportion of reference price jumps in any order book state* q *is smaller than one on* Ω / W:

$$\sup_{q \in \Omega / W} \frac{u(q) + d(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} < 1.$$

Recall that $q(n)$ represents the state of the order book after the n-th event and $p_{ref}(n)$ is the reference price after the n-th event, we consider here the process of reference price increments $c(n)$, defined as the reference price change at the n-th event:

$$c(n) = p_{ref}(n) - p_{ref}(n-1).$$

We have the following theorem on the ergodicity of the Markov chain $Y(n) = (q(n), c(n))$:

Theorem 4. *Under Assumptions 4, 5, 7, 8, 9 and 10, the embedded discrete-time Markov chain $Y(n) = (q(n), c(n))$ is V-uniformly ergodic and positive Harris recurrent.*

1.2.3 Scaling limit

Another important element in order book modeling is the scaling limit of the price process. Let

$$N(t) = \inf\{n,\ J_n \leq t\}$$

be the number of events until time t, with the convention $\inf\{\emptyset\} = 0$. Let $Z(n)$ be the cumulative price change until the n-th event, that is $Z(0) = 0$ and for $n \geq 1$:

$$Z(n) = \sum_{i=1}^{n} c(i).$$

We have

$$Z(N(t)) = p_{ref}(t) - p_{ref}(0).$$

Thus it represents the reference price at time t recentered its starting value. The following theorem shows that the rescaled price process $\hat{S}^{(n)}(t) := \frac{Z(\lfloor nt \rfloor)}{\sqrt{n}}$ in event time converges to a Brownian motion under the preceding assumptions:

Theorem 5. *Under Assumptions 4, 5, 7, 8, 9 and 10, if $\mathbb{E}_{\pi^*}[c(0)] = 0$, then the series*

$$\sigma^2 = \mathbb{E}_{\pi^*}[c_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_{\pi^*}[c_0 c_n],$$

converges absolutely, with π^* the invariant distribution of $(q(n), c(n))$. Furthermore, if $Y(0) \sim \pi^*$, we have the following convergence in law in $D[0,\infty)$:

$$\hat{S}^{(n)}(t) \overset{n\to\infty}{\to} \sigma B(t),$$

where $B(t)$ is a standard Brownian motion.

Consider now the following additional assumption:

Assumption 11. *There exists some $m > 0$, such that*

$$\inf_{q\in\Omega} \Big\{\sum_i (f_i^*(q) + g_i^*(q)) + u(q) + d(q)\Big\} > m.$$

Then we have the following result on the convergence of the rescaled reference price process in calendar time:

$$\tilde{S}^{(n)}(t) = \frac{Z\big(N(nt)\big)}{\sqrt{n}}.$$

Theorem 6. *Let τ_n be the inter-arrival time between the n-th and the $(n-1)$-th jumps of the Markov process X. Under Assumptions, 4, 5, 7, 8, 9, 10 and 11, the process $(q(n), c(n), \tau(n))$ is positive Harris recurrent. Furthermore, if $\mathbb{E}_{\pi^*}[c(0)] = 0$ and $Y(0) \sim \pi^*$, then*

$$\tilde{S}^{(n)}(t) \overset{n\to\infty}{\to} \frac{\sigma}{\sqrt{\mathbb{E}_{\pi^{**}}[\tau(1)]}} B(t),$$

with π^{**} the invariant distribution of $(q(n), c(n), \tau(n))$.

The above two theorems discuss the scaling limit of the underlying reference price process. For the more usual process such as $p_{bestbid}(t)$, $p_{bestask}(t)$ and $p_{mid}(t)$, the same result still apply since they are all bounded by $2K$ with respect to $p_{ref}(t)$.

2 Part II: Tick Value Effects

The tick value is the minimum price change imposed by the market designer on a traded asset. It is one of the most important structural parameters that affect the microstructure of the underlying asset and thus its macroscopic properties. In this part, we present our theoretical and empirical results in studying the role of tick size in high frequency trading.

2.1 The Effects of Tick Value Changes on Market Microstructure: Analysis of the 2014 Japanese Experiment

In Chapter III, we study the effects of tick value reduction for large tick assets. We aim at demonstrating that the approach introduced in Dayri and Rosenbaum (2012) allows for an *ex ante* assessment of the consequences of a tick value change on the microstructure of a large tick asset. The data of the Japanese tick value reduction pilot program are analyzed in light of this methodology. For these assets, the notion of implicit spread is introduced using the high frequency indicator η from the model with uncertainty zones of Robert and Rosenbaum (2011). This enables us to forecast the future cost of market and limit orders after a tick value change. Our results are shown to be very accurate. Furthermore, we are able to define an "optimal tick value" for each asset that helps classify the assets according to the relevance of their tick value, before and after its modification.

2.1.1 Cost of trading and high frequency price dynamics

The high frequency indicator η

We propose to use a unique high frequency indicator, the parameter η (which can be easily estimated), to summarize the high frequency features of the asset. This indicator, already used in the queue-reactive model for calibration purpose, allows us to build an estimate for the relative cost of market orders and limit orders, and is directly linked to the tick size. In the first part of this chapter, we show why this indicator is far more subtle and suitable for microstructure analysis than any other measurement, like the ones based on the conventional bid-ask spread or on the market depth.

The model with uncertainty zones assumes the existence of a latent efficient price process X_t, and that a transaction at a certain price level can happen only when the price level is close enough to the efficient price. This proximity is quantified by the parameter η: the distance between the possible transaction price and the efficient price must be smaller than $\alpha(1/2+\eta)$, with α the asset's tick value. The zone with width $2\eta\alpha$ around the mid price is called uncertainty zone. When the efficient price is inside it, both buy and sell market orders can occur.

Perceived tick size and cost of market orders

The parameter η measures the mean-reversion level of the transaction price due to the existence of the tick value. For large tick assets, η is also related to the perceived tick size of the asset by market participants: a small η (< 0.5) means that the tick value appears too large while a large η (> 0.5) means that it is considered too small, see Dayri and Rosenbaum (2012). Knowing η also enables us to compute the cost of the orders. Take for example a market order at price P_t leading to an upward price change at time t. The *e*x post expected cost of this order is given by:

$$P_t - \mathbb{E}[X_\infty | X_t] = \alpha/2 - \eta\alpha.$$

Prediction of the cost of market and limit orders

Let us consider a large tick asset with current tick value α_0 and associated high frequency indicator η_0. When the tick value is changed to α, the following prediction formula for the new value of η (and thus the cost of market and limit orders) can be established based on the invariance of the volatility with respect to the tick value:

$$\eta \sim (\eta_0 + 0.1)(\frac{\alpha_0}{\alpha})^{1/2} - 0.1. \tag{1}$$

This formula, which is valid for large tick assets ($\eta \leq 0.5$), enables us:

- To tell whether the asset will remain a large tick asset after the tick value change: if the predicted value of η is greater or equal than 0.5, the asset is predicted to become a small tick asset after the tick value change.
- To predict the new value of η: if the predicted value of η is smaller than 0.5, the above formula computes the estimated η after the tick value change.

The optimal tick value

Being able to predict the value of η after a tick value change not only helps the market designer to forecast the consequences of their measures on the market microstructure, but also paves the way for defining a notion of "optimal tick value". Although different market participants can have quite opposite views on what a good tick value is, there are in general two main objectives in a tick value change program:

- The bid-ask spread should be close to one tick, ensuring the presence of liquidity in the order book.
- Transaction costs should be close to zero for market orders. In that case, the market is efficient and market makers do not take advantage of the tick value to the detriment of final investors acting mainly as liquidity takers.

In our approach, an asset enjoys a relevant tick value if it is a large tick asset and its η parameter is close to $1/2$. Thus we have the following formula for the optimal tick value:

$$\alpha_{opt} = \frac{\alpha_0(\eta_0 + 0.1)^2}{0.36}.$$

2.1.2 Analysis of the Tokyo stock exchange pilot program on tick values

The Tokyo stock exchange pilot program, in which 55 stocks are involved in a tick size reduction plan, consists in three phases: Phase 0 (before the pilot program), from June 3, 2013 to January 13, 2014; Phase 1 (between the first and the second implementation of the tick value reduction program), from January 14, 2014 to July 21, 2014; Phase 2 (after the implementation of the second tick value reduction program), from July 22, 2014 to December 30, 2014. We assess the quality of the prediction formula on this experiment by predicting the outcomes of the first and second tick value reductions and comparing our forecasts to the empirical results.

Classification of the stocks

The 55 stocks are classified according to their conventional spread S (in ticks) and to the level of market order cost. We first split the stocks into three groups:

- Small tick stocks: $S > 1.6$.
- Large tick stocks: $S \leq 1.5$.
- Ambiguous cases: $1.5 < S \leq 1.6$.

The cost of market order being $\alpha/2 - \eta\alpha$, we use the high frequency indicator η to distinguish between balanced stocks (for which the cost of market order is close to 0) and market makers favorable stocks (where market makers obtain significant profit from liquidity takers thanks to the large tick value):

- Balanced stocks: $\eta \geq 0.4$.
- Market makers favorable stocks: $\eta < 0.4$.

A stock is considered to have a "suitable" tick value if it is both a large tick stock and a balanced stock. Note that all small tick stocks are considered as balanced stocks, but they do not have "suitable" tick value, their bid-ask spread (in ticks) being too large. Among the 55 stocks involved in the pilot program, only 5 of them are considered having a suitable tick value before the start of the program. This means that a tick value modification can be beneficial for the other 50 stocks.

Phase 0 - Phase 1

During Phase 1, 12 stocks being large tick assets in Phase 0 are involved in the tick value reduction program. These stocks are selected to test the prediction quality of Formula 1 on the new value of η in Phase 1 (η_1^p). The predicted value η_1^p tells directly whether the asset will remain a large tick asset and whether it will be balanced after the tick value modification. We use the following criteria:

- If $\eta_1^p \geq 0.55$, the asset is predicted to become a small tick asset after the tick value change.
- If $\eta_1^p < 0.5$, the asset is predicted to remain a large tick asset after the tick value change, with the forecast value for the new η being meaningful and given by η_1^p.
- We qualify the situation $0.5 \leq \eta_1^p < 0.55$ as an "ambiguous" case between large tick and small tick.

We obtain an average relative prediction error for η of 18% along with very tight confidence intervals for these 12 stocks in Phase 1. Having such accurate predictions, it is no surprise that we are also able to forecast the category an asset will belong to after the tick value modification with very high success rate (only 2 errors).

Phase 1 - Phase 2

More stocks (48) are affected in Phase 2 by the tick value reduction program. For these stocks, we conduct a similar analysis as the one for Phase 0 - Phase 1. An excellent accuracy is once again obtained for the prediction of η and the category of these stocks after the tick value modification: the average relative prediction error is reported to be less than 17%, with a success rate of 85% for the prediction of the stock's category. These results confirm the excellent prediction quality of Formula 1 and the ability of our methodology to forecast ***ex ante*** the consequences of a tick value change on the microscopic properties of the asset.

2.2 An Agent-based Model on Order Book Dynamics

In Chapter IV, we introduce a simple agent-based model on order book dynamics, which gives insights on the relationships between traded volume V, price volatility σ, tick size α, bid-ask spread ϕ and the order book equilibrium form $L(x)$ ($L(x)$ denotes the quantities of buying/selling orders between $P(t)$ and $P(t)+x$, where $P(t)$ denotes the market underlying efficient price, whose role is equivalent to the role of p_{ref} in Part I).

2.2.1 Model with infinitesimal tick size

We assume that $P(t) = P_0 + Y(t)$, with $Y(t)$ a compound Poisson process with jump rate λ^i and size (denoted by B) distribution ψ (defined on $\mathbb{R}$). We impose $\mathbb{E}[B] = 0$ so that $P(t)$ is a martingale. In such case, we have $\sigma := \frac{\mathbb{V}[P(t)]}{t} = \lambda^i \mathbb{E}[B^2]$, where the term σ represents the macroscopic volatility.

Agents

We assume that there exist three types of traders in the market:

- One informed trader: the informed trader receives the value of the price jump size B right before it happens. He then sends his trades based on this information to gain profit. We assume that he can only send market orders.

- One noise trader: the noise trader sends random market orders to the market. We assume that these trades follows a compound Poisson process, with arrival rate λ^u and volume distribution κ^u in $\mathbb{R}$ (positive volume represents a buying order, while negative volume represents a selling order).

- Market makers: the market makers receive the value of the price jump size B right after it happens. They place limit orders and try to make profit. We assume that they are risk neutral.

The following greedy assumption on the informed trader's behavior enables us to link the informed trader's trade size Q^i, the noise trader's trade size Q^u, the price jump size B and the order book's cumulative shape $L(x)$. We denote the repartition functions of B, Q^u, Q^i and Q

(the unconditional trade size) respectively by $F_\psi(x)$, $F_{\kappa^u}(x)$, $F_{\kappa^i}(x)$ and $F_\kappa(x)$, and define the inverse cumulative order book shape function $L^{-1}(q) := \underset{x}{\operatorname{argmax}}\{x|L(x) = q\}$.

Assumption 12. *Based on the received value of B and the cumulative order book shape function $L(x)$ provided by market makers, the informed trader sends his trade in a greedy way such that he wipes out all the available liquidity in the limit order book till the level $P(t) + B$. Thus, his trade size Q^i satisfies:*

$$Q^i = L(B).$$

From the above assumption, we can prove the following theorem:

Theorem 7. *The repartition functions $F_{\kappa^u}(x)$, $F_{\kappa^i}(x)$ and $F_\kappa(x)$ satisfy, with $r = \frac{\lambda_i}{\lambda_i+\lambda_u}$:*

$$\begin{aligned} F_{\kappa^i}(q) &= F_{\psi^i}(L^{-1}(q)) \\ F_\kappa(q) &= rF_{\kappa^i}(q) + (1-r)F_{\kappa^u}(q). \end{aligned}$$

The equilibrium limit order book shape

We now study the behavior of market makers under the greedy assumption. Consider the profit of passive selling orders placed between $P(t) + x$ and $P(t) + x + \delta p$ for some $x, \delta p > 0$, given the fact that a buying transaction of size Q happens with $Q \geq L(x + \delta p)$, that is all these orders are executed. If the transaction comes from the informed trader, market makers who place these limit orders lose money due to the greedy assumption, and they gain profit if the transaction comes from the noise trader as the difference between the average executed price and the efficient price $P(t)$. We denote the conditional expected *ex post* gain of these orders by $G(x, \delta p)$, and define the average profit per unit at x, denoted by $G(x)$, as the limit of the average *ex post* profit of the passive orders placed between $P(t) + x$ and $P(t) + x + \delta p$:

$$G(x) = \lim_{\delta p \to 0^+} \frac{G(x, \delta p)}{L(x + \delta p) - L(x)}.$$

A positive $G(x)$ means that new limit orders at x are still profitable. As market makers are assumed to be risk neutral, it is natural to make the following zero profit assumption (with $l(x) := L'(x)$):

Assumption 13. *For every $x \in \mathbb{R}^+$, if $L(x) > 0$ and $l(x) > 0$, then the conditional expected average gain per unity of passive orders placed at x, given the fact that they are totally executed, is equal to 0, that is, when $L(x) > 0$ and $l(x) > 0$:*

$$G(x) = 0.$$

Equivalently, we have:

$$x = r\frac{\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - F_\kappa(L(x))}.$$

Unlike the traditional zero-profit assumption (such as that used in Glosten and Milgrom (1985)) which assumes that market makers make no profit at all, under the above assumption, “fast” market makers can still make profit from their limit orders placed before the equilibrium is attained due to the competition among them. This point will be made clearer when the tick value constraint is added.

The following theorem shows that the bid-ask spread emerges naturally from the above assumptions as well as an equilibrium cumulative order book shape:

Theorem 8. *The cumulative limit order book shape function is uniquely determined. We have, $L(x) = 0$, for $x \in [-\eta,\eta]$, where η is the unique solution of the following equation:*

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\eta},1)].$$

Furthermore, for $x > \eta$:

$$L(x) = F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x},1)]),$$

and for $x < -\eta$:

$$L(x) = -F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{-x},1)]).$$

In particular, the intrinsic bid-ask spread satisfies $\phi = 2\eta$.

Actually, the existence of the informed trader prevents market makers from posting their limit orders too close to the efficient price: the expected profit of these orders is negative as they are very vulnerable to large price jumps. In our model, market makers can only make profit from the noise trades. The minimum distance η to which their limit orders start to make profit depends naturally on the proportion of noise trades as well as on the distribution of the price jump.

2.2.2 Tick size

Constrained bid-ask spread

When we constrain price changes by the tick size, the cumulative order book $L(x)$ becomes a piece-wise constant function. We denote by $G^d(i)$ the expected average gain of passive orders placed at the i-th limit given the fact that a transaction happens and they are completely executed. We make a similar zero profit assumption in such case. We write $d := \bar{P}(t) - P(t)$, with $\bar{P}(t)$ the smallest possible price level that is greater or equal to the efficient price $P(t)$, and $l^d(i)$ the number of limit orders placed at the i-th limit:

Assumption 14. *For every $i \in \mathbb{N}^+$, if $l^d(i) > 0$, then the conditional expected average gain of the passive orders placed at $d + (i-1)\alpha$, given the fact that they are totally executed, is equal to 0, that is, when $l^d(i) > 0$:*

$$G^d(i) = 0.$$

Equivalently, we have:

$$d + (i-1)\alpha = \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(i-1)\alpha}]}{1 - F_\kappa(L^d(d+(i-1)\alpha))}.$$

When one limit is empty, its potential expected gain can be defined as the average gain/loss of an infinitesimal passive order placed at this limit, under the condition that it is executed completely. When the potential expected gain is positive, market makers naturally place new passive orders as these orders are profitable in expectation. This idea gives the following assumption:

Assumption 15. *For every $i \in \mathbb{N}^+$, if $l^d(i) = 0$, then the potential conditional expected average gain of the passive orders placed at $d + (i-1)\alpha$, given the fact that they are totally executed, is less than or equal to 0, that is, when $l^d(i) = 0$:*

$$G^d(i) \leq 0.$$

Equivalently, we have:

$$d+(i-1)\alpha \leq \frac{r\mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]}{1-rF_{\psi}(d+(i-1)\alpha)-(1-r)F_{\kappa^u}(\max(0,L^d(d+(i-2)\alpha)))}.$$

One interesting theorem can be obtained from these assumptions. It suggests that the average bid-ask spread, which is now constrained by the tick size, is a linearly increasing function of the tick size α:

Theorem 9. *The average bid-ask spread ϕ_α satisfies the following equation:*

$$\phi_\alpha = \alpha + \phi,$$

where ϕ is the intrinsic bid-ask spread of the asset when the tick value is equal to 0.

The cumulative limit order book shape is shown to satisfy the following theorem:

Theorem 10. *The cumulative limit order book shape function is uniquely determined. We have, $l^d(i) = 0$ for all $-k_l^d < i < k_r^d$, where k_l^d and k_r^d are two positive integers determined by the following equations:*

$$\begin{aligned} k_r^d &= 1+\lceil\frac{\eta-d}{\alpha}\rceil, \\ k_l^d &= \lceil\frac{\eta+d}{\alpha}\rceil, \end{aligned}$$

where η is the unique solution of the following equation:

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\eta},1)].$$

For $h \geq k_r^d$:

$$L^d(d+(h-1)\alpha) = F_{\kappa^u}^{-1}(\frac{1}{1-r}-\frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+(h-1)\alpha},1)]),$$

for $h \leq -k_l^d$:

$$L^d(d+h\alpha) = -F_{\kappa^u}^{-1}(\frac{1}{1-r}-\frac{r}{1-r}\mathbb{E}[\max(\frac{B}{-d-h\alpha},1)]).$$

Priority value

The priority value, already discussed in the queue-reactive model to explain the concave cancellation rate, can be formulated in this model as the difference between the expected profit of the order placed on top and that placed at the bottom of the queue:

Theorem 11. *The priority value at the $i-$th limit can be written as:*
For $i = k_r^d$,

$$\begin{aligned} \tilde{G}^d(i) = \mathbb{E}[B\mathbf{1}_{B\geq d+(k_r^d-1)\alpha}]\{&\frac{1}{\mathbb{E}[\max(\frac{B}{d+(k_r^d-1)\alpha},1)]-F_\psi(d+(k_r^d-1)\alpha)} \\ &-\frac{1}{\frac{1+r}{2r}-F_\psi(d+(k_r^d-1)\alpha)}\}, \end{aligned}$$

for $i > k_r^d$,

$$G_p^d(i) = \mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]\{\frac{1}{\mathbb{E}[\max(\frac{B}{d+(i-1)\alpha},1)] - F_\psi(d+(i-1)\alpha)} - \frac{1}{\mathbb{E}[\max(\frac{B}{d+(i-2)\alpha},1)] - F_\psi(d+(i-1)\alpha)}\}.$$

The above formulas can be easily generalized to compute the priority value of an order placed at any position in the queue.

2.2.3 Information propagation

We then discuss the case when the market makers no longer hold exact information on the price $P(t)$, and add a minimum reaction time between the moment when the informed trader receives the updated information and the moment when he sends orders to take the liquidity. We assume here that $\alpha = 0$ in order to simplify our analysis.

The main question faced by market makers when they no longer know the value of $P(t)$ is whether they should refill the gaps in the order book once the liquidity has been taken by a market order. If the market order is sent by the noise trader, then market makers should send limit orders to refill the limit order book. But if the market order comes from the informed trader, refilling the order book leads to further losses to market makers as the newly inserted orders will probably again be consumed by the informed trader.

For a given market order of volume q (until the price level p') at the moment t, write $P(t-)$ the efficient price before this trade, which we assume is known to market makers. Immediately after this trade, the gain of a limit order placed between $P(t-)+x$ and $P(t-)+x+\delta p$ can be written as (v^i is a random variable with $v^i = 1$ if the last trade comes from the informed trader, $v^i = 0$ if the last trade comes from the noise trader):

$$g^{p'}(x,\delta p) = v^i \tilde{g}^{p'}(x,\delta p) + (1-v^i)g(x,\delta p),$$

with $\tilde{g}^{p'}(x,\delta p)$ representing the gain in the case of an informed transaction, and $g(x,\delta p)$ that in the case of a noise transaction. Denote by $L^-(p')$ the cumulative order book quantity right before the trade at the price level $P(t^-)+p'$, $r^{p'} = \frac{r(1-F_\psi(p'))}{1-rF_\psi(p')-(1-r)F_{\kappa^u}(L^-(p'))}$ and the best ask price $a^{p'}$ (we consider a buy market order) right after the trade, we have the following theorem concerning the equilibrium order book state immediately after this trade:

Theorem 12. *Assume the reaction rate λ^a is much larger than the information arrival rate λ^i and noise trade arrival rate λ^u, that is:*

$$\begin{aligned} r^i &= \frac{\lambda^i}{\lambda^i+\lambda^u+\lambda^a} \approx 0 \\ r^u &= \frac{\lambda^u}{\lambda^i+\lambda^u+\lambda^a} \approx 0. \end{aligned} \tag{2}$$

The best ask price $a^{p'}$ lies in the interval $(r^{p'}, p')$ and satisfies the following approximate equation:

$$\frac{1+r}{2r} \quad \approx \quad (1-r^{p'})\frac{\mathbb{E}[B\mathbf{1}_{B>a^{p'}}]}{a^{p'}-r^{p'}p'}+F_{\psi}(a^{p'}).$$

Moreover, the equilibrium order book state $L(x)$ immediately after a transaction satisfies:

for $x<p'$:

$$L(x) \quad \approx \quad F_{\kappa^u}^{-1}(\frac{1}{1-r}-\frac{r(1-r^{p'})\mathbb{E}[\max(B,x)]-rr^{p'}(p'-x)F_{\psi}(x)}{(1-r)(x-r^{p'}p')}),$$

and for $x \geq p'$:

$$G^{p'}(x) \quad = \quad x-r^{p'}p'-\frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1-rF_{\psi}(x)-(1-r)F_{\kappa^u}(L(x))}-\frac{r^{p'}r\mathbb{E}[B\mathbf{1}_{B>x-p'}]}{1-rF_{\psi}(x-p')-(1-r)F_{\kappa^u}(L(x))}.$$

In particular, at the trade price p', we have:

$$L(p') \quad \approx \quad F_{\kappa^u}^{-1}(\frac{1}{1-r}-\frac{r}{1-r}\mathbb{E}[\max(\frac{B}{p'},1)])=L^{-}(p').$$

The above theorem gives lower and upper bounds on the new best bid/ask limit after a new transaction, as well as a description of the order book state around the transaction price p'. In our setting, the cumulative number of orders stays almost the same at p' immediately after the transaction. The new best bid/ask position depends on the volume of the last trade, but is strictly smaller than the transaction price (in the buying case).

Limit order book recovery speed

If the last trade is indeed an informed transaction, orders that are placed before p' are exposed to the risk of being overrun once more by the informed trader. However, as the time advances, if no further trades arrive to take away these newly inserted orders, market makers may adjust their expected potential gain. Then the limit order book state will recover gradually to its original equilibrium state. These intuitions are formalized in the following theorem.

Let $G^{p'}(x,\delta t)$ denotes the conditional expected gain per unity at the price $P^{-}(t)+x$ at the moment $t+\delta t$, given the fact that no trade has arrived in the market between t and $t+\delta t$. We write $G^{p'}(x)$ for the conditional expected gain per unity given the fact that the last trade is issued by an informed trader, and $G(x)$ represents the standard conditional expected average gain per unity. We have the following result:

Theorem 13. *Given the fact that no trade arrives between $(t,t+\delta t]$. We have, for $x<p'$,*

$$G^{p'}(x,\delta t)=r_{\delta t}^{p'}\tilde{G}^{p'}(x)+(1-r_{\delta t}^{p'})G(x),$$

with

$$r_{\delta t}^{p'} \quad = \quad \frac{1}{1+\frac{1-r^{p'}}{r^{p'}}e^{\lambda^a\delta t}}.$$

When $\delta t \to \infty$, $r_{\delta t}^{p'} \to 0$, *and* $G^{p'}(x, \delta t) \to G(x)$, *that is, if the trade is intiated by the noise trader, the limit order book shape gradually recovers after this trade to its stationary state, with approximative speed* $\sim e^{-\lambda^a t}$.

Part I

Limit Order Book Modelling

CHAPTER I

The queue-reactive model

Abstract

Through the analysis of a dataset of ultra high frequency order book updates, we introduce a model which accommodates the empirical properties of the full order book together with the stylized facts of lower frequency financial data. To do so, we split the time interval of interest into periods in which a well chosen reference price, typically the midprice, remains constant. Within these periods, we view the limit order book as a Markov queuing system. Indeed, we assume that the intensities of the order flows only depend on the current state of the order book. We establish the limiting behavior of this model and estimate its parameters from market data. Then, in order to design a relevant model for the whole period of interest, we use a stochastic mechanism that allows to switch from one period of constant reference price to another. Beyond enabling to reproduce accurately the behavior of market data, we show that our framework can be very useful for practitioners, notably as a market simulator or as a tool for the transaction cost analysis of complex trading algorithms.

Keywords: Limit order book, market microstructure, high frequency data, queuing model, jump Markov process, ergodic properties, volatility, mechanical volatility, market simulator, execution probability, transaction costs analysis, market impact.

1 Introduction

Electronic limit order books (LOB for short), where market participants send their buy and sell orders via a continuous-time double auction system, are nowadays the dominant mode of exchange on financial markets. Consequently, understanding the LOB dynamics has become a fundamental issue. Indeed, a deep knowledge of the LOB's behavior enables policy makers to design relevant regulations, market makers to provide liquidity at cheaper prices, and investors to save transaction costs while mounting and unwinding their positions, thus reducing the cost of capital of listed companies. Furthermore, it can also provide insights on the macroscopic features of the price which emerges from the LOB.

In the seminal work on zero intelligence LOB models of Smith, Farmer, Gillemot, and Krishnamurthy (2003), a mean-field approach is suggested in order to study the properties of the LOB. In such models, the underlying assumption is that the order flows follow independent Poisson processes. Although this hypothesis is not really compatible with empirical observations, the authors show that its simplicity allows for the derivation of many interesting formulas, some of them being testable on market data. This work has been followed by numerous developments. For example, in Cont, Stoikov, and Talreja (2010), the probabilities of various order book related events are computed in this framework, whereas stability conditions of the system are studied in Abergel and Jedidi (2011). We wish to extend this approach in two directions. On the one hand, we want our model to be more consistent with market data, so that we can give new insights on the dynamics of the LOB. On the other hand, we aim at providing a useful and relevant tool for market practitioners, notably in the perspective of transaction costs analysis.

Under the first in first out rule (which we assume in the sequel), a LOB can be considered as a high-dimensional queuing system, where orders arrive and depart randomly. We consider the three following types of orders:

- Limit orders: insertion of a new order in the LOB (a buy order at a lower price than the best ask price, or a sell order at a higher price than the best bid price).
- Cancellation orders: cancellation of an already existing order in the LOB.
- Market orders: consumption of available liquidity (a buy or sell order at the best available price).

In practice, market participants (or their algorithms) analyze many quantities before sending a given order at a given level. One of the most important variables in this decision process is probably the distance between their target price and their "reference market price", typically the midprice. This reference price is linked with the order flows since it is usually determined by the LOB state. This interconnection makes the design of LOB models quite intricate. To overcome this difficulty, we split the time interval of interest into periods of constant reference price, and consider two parts in our modeling. First, we study the LOB as a Markov queuing system during the time periods when the reference price is constant. Then, we investigate the dynamics of the reference price. Such a framework is particularly suitable for large tick assets[1], for which constant reference price periods are quite long and allow for accurate parameter estimations.

Two kinds of public information are available to market participants at the high frequency scale: the historical order flows and the current state of the LOB. In this paper, we are mostly interested in how the state of the LOB impacts market participants decisions. Surprisingly enough, this question has been rarely considered in the literature. Let us mention as an exception the interesting approach in Gareche, Disdier, Kockelkoren, and Bouchaud (2013), where the impact of the LOB state on the queue dynamics is analyzed through PDE type arguments. Within periods of constant reference price, we model the LOB as a continuous-time Markov jump process, and estimate its infinitesimal generator matrix under various assumptions on the information set used by market participants. From these results, we are able to analyze how market participants react towards different configurations of the LOB. Furthermore, we provide the asymptotic distributions of the LOB. The level of realism of our approaches is assessed by comparing expected features from the models with empirical ones. Thus, all our developments are illustrated on two specific examples of large tick stocks on Euronext Paris: France Telecom and Alcatel-Lucent (in appendix).

In the second part of the paper, we extend our framework by allowing reference price moves, so that our model also accommodates macroscopic properties of the asset (roughly summarized by the volatility). Modifications of the reference price[2] will possibly occur provided one of the best queues is totally depleted or a new order is inserted within the spread. This model is called

[1]A large tick asset is defined as an asset whose bid-ask spread is almost always equal to one tick, see Dayri and Rosenbaum (2012). In practice, our framework can be considered relevant for any asset whose average spread is smaller than 2.5 ticks.

[2]Note that the reference price will not be exactly the midprice, see Section 2.2.

“queue-reactive model". In particular, it enables us to bring to light a quantity, the “maximal mechanical volatility", which represents the amount of price volatility generated by the generic randomness of the order flows. In practice, this parameter is typically smaller than the empirical volatility estimated from market data. The reason for this is simple: the market does not evolve like a closed physical system, where the only source of randomness would be the endogenous interactions between participants. It is also subject to external informations, such as the news, which increase the volatility of the price. Hence, it will be necessary to introduce an exogenous component within the queue-reactive model.

Throughout the paper, we illustrate the fact that many useful short term predictions can be computed in our framework: execution probabilities of passive orders, probability of price increase... More importantly, we show that the queue-reactive model turns out to be a very relevant market simulator, notably in view of the analysis of complex trading tactics, using for example a mixture of market and limit orders.

The paper is organized as follows. In Section 2, we consider periods when the reference price is constant. We first present a very general framework for the LOB dynamics and then introduce three specific models. The first one is a birth and death process in which the queues are assumed to be independent. In this setting, we are able to fully characterize the asymptotic behavior of the LOB. The second approach is a queuing system in which the bid and ask sides are independent, but the first two lines on each side can exhibit correlations. We show that this model can be seen as a quasi birth and death process (QBD for short) and thus admits a matrix geometric solution as its invariant distribution. In the last approach, we allow for cross dependences between bid and ask queues. An application of these models to the computation of execution probabilities is presented at the end of the same section. In Section 3, we investigate the dynamics of the reference price. In particular, we build the queue-reactive model which is a relevant LOB model for the whole time period of interest. We end this section by showing how our framework can be used for transaction costs and market impact analysis of high frequency trading strategies. A conclusion and some perspectives are given in Section 4. Some proofs and further empirical results are gathered in an appendix.

2 Dynamics of the LOB in a period of constant reference price

Within time periods when the reference price is constant, we consider three different models for the LOB. These models can be jointly introduced through the general framework we present now.

2.1 General Framework

In the general framework, the LOB is seen as a $2K$-dimensional vector, where K denotes the number of available limits on each side[3], see Figure I.1. The reference price p_{ref} defines the center of the $2K$-dimensional vector, and divides the LOB into two parts: the bid side $[Q_{-i} : i = 1,...,K]$ and the ask side $[Q_i : i = 1,...,K]$, where $Q_{\pm i}$[4] represents the limit at the distance $i - 0.5$ ticks to the right $(+i)$ or to the left $(-i)$ of p_{ref}. The number of orders at Q_i is denoted by q_i. We assume that on the bid (resp. ask) side, market participants send buy (resp. sell) limit orders, cancel existing buy (resp. sell) orders and send sell (resp. buy) market orders. We consider a constant order size at each limit. However, the order sizes at the different limits are allowed to

[3]Note that an empty limit can be part of the LOB in our setting.

[4]To simplify our notations, we write $*_i/*_{-i}$ as $*_{\pm i}$, and $*_{-i}/*_i$ as $*_{\mp i}$.

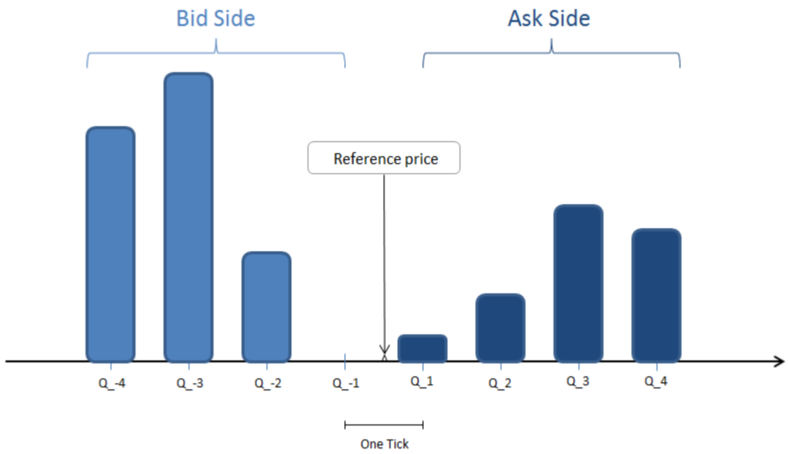

Figure I.1: Limit order book

be different. In practice, these sizes can be chosen as the average event sizes observed at each limit Q_i (AES_i for short)[5].

The $2K$-dimensional process $X(t) = (q_{-K}(t), \ldots, q_{-1}(t), q_1(t), \ldots, q_K(t))$ is then modeled as a continuous-time Markov jump process in the countable state space $\Omega = \mathbb{N}^{2K}$, with jump size equal to one. For $q = (q_{-K}, \ldots, q_{-1}, q_1, \ldots, q_K) \in \Omega$, and $e_i = (a_{-K}, \ldots, a_i, \ldots, a_K)$, where $a_j = 0$ for $j \neq i$ and $a_i = 1$, the components $\mathcal{Q}_{q,p}$ of the infinitesimal generator matrix $\mathcal{Q}$ of the process $X(t)$ are assumed to be of the following form:

$$\begin{aligned}
\mathcal{Q}_{q,q+e_i} &= f_i(q) \\
\mathcal{Q}_{q,q-e_i} &= g_i(q) \\
\mathcal{Q}_{q,q} &= -\sum_{p\in\Omega, p\neq q} \mathcal{Q}_{q,p} \\
\mathcal{Q}_{q,p} &= 0, \text{otherwise.}
\end{aligned}$$

We now give a theoretical result on the ergodicity of the system under two very general assumptions. Let us denote by $P_{q,p}(t)$ the transition probability from state q to state p in a time t. Recall that a Markov process in a countable state space is said to be ergodic if there exists a probability measure π that satisfies $\pi P = \pi$ (π is called invariant measure) and for every q and p:

$$\lim_{t\to\infty} P_{q,p}(t) = \pi_p.$$

We consider the two following assumptions.

Assumption 16. *(Negative individual drift) There exist a positive integer C_{bound} and $\delta > 0$, such that for all i and all $q \in \Omega$, if $q_i > C_{bound}$,*

$$f_i(q) - g_i(q) < -\delta.$$

[5]In our framework, AES_i is a more suitable choice than ATS (Average Trade Size) that computes only the average size of market orders, see Section 5.6 in appendix for more details.

Assumption 17. *(Bound on the incoming flow) There exists a positive number H such that for any $q \in \Omega$,*

$$\sum_{i \in [-K,\dots,-1,1,\dots,K]} f_i(q) \le H.$$

Assumption 16 can be interpreted as follows: the queue size of a limit tends to decrease when it becomes too large. Assumption 17 ensures no explosion in the system: the order arrival speed stays bounded for any given state of the LOB. Under these two assumptions, we have the following ergodicity result for the $2K$-dimensional queuing system. The proof is given in appendix.

Theorem 1. *Under Assumptions 16 and 17, the $2K$-dimensional Markov jump process $X(t)$ is ergodic.*

This theorem is the basis for the asymptotic study of the LOB dynamics in the following sections.

2.2 Data description and estimation of the reference price

2.2.1 The database

The data used in our empirical studies are collected from Cheuvreux's[6] LOB database, from January 2010 to March 2012, on Euronext Paris. It records the LOB data (prices, volume and number of orders) up to the fifth best limit on both sides, whenever the LOB state changes. Note that we remove market data corresponding to the first and last hour of trading, as these periods have usually specific features because of the opening/closing auction phases. Two large tick European stocks, France Telecom and Alcatel-Lucent, are studied and they exhibit very similar behaviors. Some characteristics of these two stocks are given in Table 1. We have chosen the stock France Telecom as illustration example for all the developments in the sequel. The results for Alcatel-Lucent can be found in appendix. Although only stocks are considered in this paper, our method applies also to other financial assets, such as interest rates or index futures (among which large tick assets are quite numerous, see Dayri and Rosenbaum (2012)).

stock	average number of orders per day	average number of trades per day	average spread size (in number of ticks)
France Telecom	159250	7282	1.43
Alcatel Lucent	129400	8626	1.99

Table I.1: Data description

2.2.2 Estimation of the reference price

As mentioned in the introduction, the estimation of a relevant reference price p_{ref} is the basis for defining the limits in the order book. Indeed, p_{ref} provides the center point of the LOB and thus the positions of the $2K$ limits. In our framework, if we write p_i for the price level of the limit Q_i, $i = -K, \dots, 1, 1, \dots, K$, we must have

$$p_{ref} = \frac{p_1 + p_{-1}}{2}.$$

[6]Cheuvreux is a brokerage firm based in Paris, formerly a subsidiary of Crédit Agricole Corporate Investment Bank, and now merged with Kepler Capital Market.

When the observed bid-ask spread is equal to one tick, p_{ref} is obviously taken as the midprice (denoted by p_{mid}) and both Q_1 and Q_{-1} are non empty. When it is larger than one tick, several choices are possible for p_{ref}. We build p_{ref} from the data the following way: when the spread is odd (in tick unit), it is still natural to use p_{mid} as the LOB center:

$$p_{ref} = p_{mid} = \frac{(p_{bestbid} + p_{bestask})}{2}.$$

When it is even, p_{mid} is no longer appropriate since it is now itself a possible position for order arrivals. In such case, we use either

$$p_{mid} + \frac{\text{tick size}}{2} \text{ or } p_{mid} - \frac{\text{tick size}}{2},$$

choosing the one which is the closest to the previous value of p_{ref}. Note that more complex methods could be used for the estimation of p_{ref}, see for example Delattre, Robert, and Rosenbaum (2013).

2.3 Model I: Collection of independent queues

We now give a first simple LOB model around a fixed reference price.

2.3.1 Description of the model

In this model, we assume independence between the flows arriving at different limits in the LOB. Three types of orders are considered: limit orders, cancellations and market orders. We suppose that the intensities of these point processes at different limits are only functions of the target queue size (that is the available volume at the considered limit Q_i). Furthermore, at a given limit, conditional on the LOB state, the arrival processes of the three types of orders are taken independent. The values of these intensities are denoted by $\lambda_i^L(n)$ (limit orders), $\lambda_i^C(n)$ (cancellations) and $\lambda_i^M(n)$ (market orders) when $q_i = n$. Moreover, the intensity functions at Q_i and Q_{-i} are chosen identical, considering the symmetry property of the LOB. We then have $\lambda_i^L(n) = \lambda_{-i}^L(n), \lambda_i^C(n) = \lambda_{-i}^C(n), \lambda_i^M(n) = \lambda_{-i}^M(n)$, and

$$\begin{aligned} f_i(q) &= \lambda_i^L(q_i) \\ g_i(q) &= \lambda_i^C(q_i) + \lambda_i^M(q_i). \end{aligned}$$

In this model, market orders sent to Q_i consume directly the volume available at Q_i. Therefore, we can have a market order at the second limit while the first limit is not empty. However, for large tick assets, this assumption is reasonable as their market order flow is almost fully concentrated on first limits ($Q_{\pm 1}$) and the estimated intensities of this flow at ($Q_{\pm i}$), $i \neq 1$ are very small. Under these assumptions, the LOB becomes a collection of $2K$ independent queues, each of them being a birth and death process.

2.3.2 Empirical study: Collection of independent queues

In Model I, the intensities of the different queues can be estimated separately. The value of K is set to 3, as our numerical experiments show that for the considered stocks, both the dynamics and empirical distributions at $Q_{\pm i}, i = 4, 5$ are quite similar to that at $Q_{\pm 3}$. This value of K will also apply to other experiments in the paper.

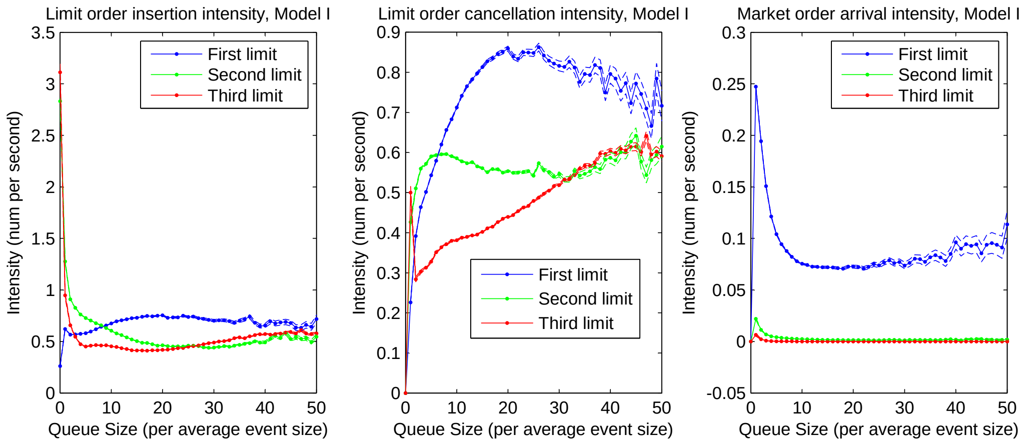

Figure I.2: Intensities at $Q_{\pm i}$, $i = 1, 2, 3$, France Telecom

The estimation method goes as follows. We define an "event" ω as any modification of the queue size. For queue Q_i, we record the waiting time $\Delta t_i(\omega)$ (in number of seconds) between the event ω and the preceding event at Q_i, the type of the event $\mathcal{T}_i(\omega)$ and the queue size $q_i(\omega)$ before the event. The queue size is then approximated by the smallest integer that is larger than or equal to the volume available at the queue, divided by the stock's average event size AES_i at the corresponding queue. We set the "type" of the event ω the following way:

- $\mathcal{T}_i(\omega) \in \mathcal{E}^+$ for limit order insertion at Q_i,
- $\mathcal{T}_i(\omega) \in \mathcal{E}^-$ for limit order cancellation at Q_i,
- $\mathcal{T}_i(\omega) \in \mathcal{E}^t$ for market order at Q_i.

When the reference price changes, we restart the recording process. Once we have collected $(\Delta t_i(\omega), \mathcal{T}_i(\omega), q_i(\omega))$ from historical data, it is easy to estimate $\lambda_i^L(n)$, $\lambda_i^C(n)$ and $\lambda_i^M(n)$ by the maximum likelihood method:

$$
\begin{aligned}
\hat{\Lambda}_i(n) &= \left(\text{mean}(\Delta t_i(\omega) | q_i(\omega) = n)\right)^{-1} \\
\hat{\lambda}_i^L(n) &= \hat{\Lambda}_i(n) \frac{\#\{\mathcal{T}_i(\omega) \in \mathcal{E}^+, q_i(\omega) = n\}}{\#\{q_i(\omega) = n\}} \\
\hat{\lambda}_i^C(n) &= \hat{\Lambda}_i(n) \frac{\#\{\mathcal{T}_i(\omega) \in \mathcal{E}^-, q_i(\omega) = n\}}{\#\{q_i(\omega) = n\}} \\
\hat{\lambda}_i^M(n) &= \hat{\Lambda}_i(n) \frac{\#\{\mathcal{T}_i(\omega) \in \mathcal{E}^t, q_i(\omega) = n\}}{\#\{q_i(\omega) = n\}},
\end{aligned}
$$

where "mean" denotes the empirical mean and $\#A$ the cardinality of the set A.

In Figure I.2, we present the estimated intensities. Data at Q_i and Q_{-i} are aggregated together (simply by combining the two collected samples) and confidence intervals (dotted lines) are computed using central limit approximations detailed in appendix. We now comment the obtained graphs.

Behaviors under the independence assumption

- Limit order insertion:

  $Q_{\pm 1}$: The intensity of the limit order insertion process is approximately a constant function of the queue size, with a significantly smaller value at 0. Note that inserting a limit order in an empty queue creates a new best limit and the market participant placing this order is the only one standing at this price level. Such action is often risky. Indeed, when the spread is different from one tick, one is quite uncertain about the position of the so-called “efficient" or “fair" price, see for example Delattre, Robert, and Rosenbaum (2013) for discussions on this notion. This smaller value can also be due to temporary realizations of the structural relation between the bid-ask spread and the volatility: if the spread is large because the inventory risk of market makers is high, the probability that anyone inserts a limit order in the spread is likely to be low, see among others Madhavan, Richardson, and Roomans (1997), Avellaneda and Stoikov (2008), Wyart, Bouchaud, Kockelkoren, Potters, and Vettorazzo (2008) and Dayri and Rosenbaum (2012) for more details about market making and the relation between spread, volatility and inventory risk.

  $Q_{\pm 2}$: The intensity is now approximately a decreasing function of the queue size. This interesting result probably reveals a quite common strategy used in practice: posting orders at the second limit when the corresponding queue size is small to seize priority. More details on this strategy are given in Section 2.4.2.

  $Q_{\pm 3}$: The intensity function shows similar properties to that at the second limit.

- Limit order cancellation:

  $Q_{\pm 1}$: The rate of order cancellation is an increasing concave function for $q_{\pm 1}$ between 0 and 25, and becomes flat/slightly decreasing for larger values. This result is in contrast to the classical way to model this flow, where one often considers a linearly increasing cancellation rate, see for example Cont, Stoikov, and Talreja (2010). On this first in first out market, the priority value, that is the advantage of a limit order compared with another limit order standing at the rear of the same queue, can be one of the reasons for this behavior. Indeed, the priority value is an increasing function of the queue size and orders having a high priority value are less likely to be canceled.

  $Q_{\pm 2}$: The rate of order cancellation attains more rapidly its asymptotic value, which is lower than for $Q_{\pm 1}$. Compared to the first limit case, market participants at the second limit have even stronger intention not to cancel their orders when the queue size increases. This is probably due to the fact that these orders are less exposed to short term market trends than those posted at $Q_{\pm 1}$ (since they are covered by the volume standing at $Q_{\pm 1}$ and their price level is farther away from the reference price).

  $Q_{\pm 3}$: The priority value is smaller at the third limit since it takes longer time for $Q_{\pm 3}$ to become the best quote if it does. The rate of order cancellation increases almost linearly for queue sizes larger than 3 AES_3. We also find a quite large cancellation rate when the queue size is equal to one, which shows that market participants cancel their orders more quickly when they find themselves alone in the queue.

- Market orders:

$Q_{\pm 1}$: The rate decreases exponentially with the available volume at $Q_{\pm 1}$. This phenomena is easily explained by market participants "rushing for liquidity" when liquidity is rare, and "waiting for better price" when liquidity is abundant.

$Q_{\pm 2}$: In practice, market orders can arrive at $Q_{\pm 2}$ only if $Q_{\pm 1} = 0$ (that is when $Q_{\pm 2}$ is the best offer queue). The shape of the intensity is very similar to the one obtained in the case of $Q_{\pm 1}$. The values are of course much smaller.

$Q_{\pm 3}$: In some rare cases, one can still find some market orders arriving at $Q_{\pm 3}$ (market orders occurring when the spread is large). The intensity function remains exponentially decreasing.

2.3.3 Asymptotic behavior under Model I

The invariant distribution of the LOB can be computed explicitly in Model I. We denote by π_i the stationary distribution of the limit Q_i, and define the arrival/departure ratio vector ρ_i by

$$\rho_i(n) = \frac{\lambda_i^L(n)}{(\lambda_i^C(n+1) + \lambda_i^M(n+1))}.$$

Then the following result for the invariant distribution is easily obtained, see for example Gross and Harris (1998):

$$\begin{aligned} \pi_i(n) &= \pi_i(0) \prod_{j=1}^{n} \rho_i(j-1) \\ \pi_i(0) &= \left(1 + \sum_{n=1}^{\infty} \prod_{j=1}^{n} \rho_i(j-1)\right)^{-1}. \end{aligned}$$

Hence the long term behavior of the LOB is completely determined by ρ. This implies that two assets can have very different flow dynamics, but still the same invariant distribution provided that their arrival/departure ratios are the same.

We now compare the asymptotic results of the model with the empirical distributions observed at $Q_{\pm 1}$, $Q_{\pm 2}$ and $Q_{\pm 3}$. To compute these empirical laws, we use a sampling frequency of 30 seconds (every 30 seconds, we look at the LOB and record its state)[7]. The results are gathered in Figure I.3, as well as the invariant distributions from a Poisson model (constant limit/market order arrival rate, linear cancellation rate, parameters estimated from the same dataset).

One can see that the invariant distributions approximate very well the empirical distributions of the LOB. This shows that in order to explain the shape of the LOB, such mean-field type approach, where the LOB profile arises from interactions between the average behaviors of market participants, can be very relevant.

2.4 Model II: Dependent case

We now present some extensions of Model I. We assume here that buy/sell market orders consume volume at the best quote limits, defined as the first non empty ask/bid queue. Thus, we consider a buy market order process with intensity λ_{buy}^M and a sell market order process with

[7]Other sampling frequencies have also been tested and the estimated distributions are found to be very similar. These sampled data will also be used to estimate the joint distributions of the LOB limits in Model II^a and II^b.

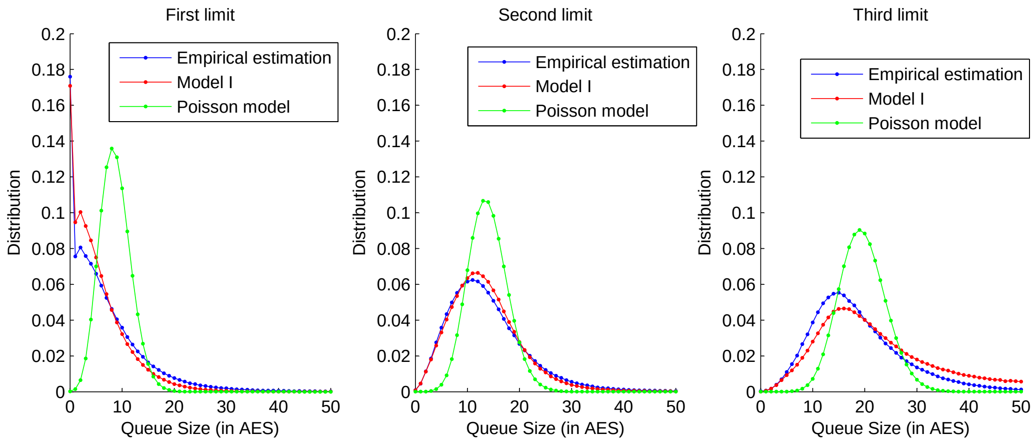

Figure I.3: Model I, invariant distributions of $q_{\pm 1}, q_{\pm 2}, q_{\pm 3}$, France Telecom

intensity λ^M_{sell}. The limit order, cancellation, and market order arrival processes are assumed to be independent conditional on the LOB state. So we can write $f_i(q)$ and $g_i(q)$ in the following form:

$$\begin{aligned} f_i(q) &= \lambda^L_i(q) \\ g_i(q) &= \lambda^C_i(q) + \lambda^M_{buy}(q)\mathbf{1}_{bestask(q)=i}, \text{if } i > 0 \\ g_i(q) &= \lambda^C_i(q) + \lambda^M_{sell}(q)\mathbf{1}_{bestbid(q)=i}, \text{if } i < 0. \end{aligned}$$

As for Model I, we consider some bid-ask symmetry, that is, for $q = [q_{-3}, q_{-2}, q_{-1}, q_1, q_2, q_3]$, $q' = [q_3, q_2, q_1, q_{-1}, q_{-2}, q_{-3}]$ and $i = 1,2,3$, $\lambda^L_i(q) = \lambda^L_{-i}(q')$, $\lambda^M_i(q) = \lambda^M_{-i}(q')$ and $\lambda^M_{buy}(q) = \lambda^M_{sell}(q')$.

2.4.1 Model II^a: Two sets of dependent queues

Institutional traders and brokers tend to place most of their limit orders at best limits, while many market makers, arbitragers and other high frequency traders stand also in queues beyond these best limits. This suggests for example that the dynamics at $Q_{\pm 2}$ may not only depend on $q_{\pm 2}$, but also on whether or not $Q_{\pm 1}$ is empty. We thus propose to use the following intensity functions for the queue $Q_{\pm 2}$: in this model, $\lambda^L_{\pm 2}$ and $\lambda^C_{\pm 2}$ are functions of $q_{\pm 2}$ and $\mathbf{1}_{q_{\pm 1}>0}$. Intensities at $Q_i, i \neq \pm 2$ remain functions of q_i only. For large tick assets, the probability that $Q_{\pm i}, i \geq 3$ is the best limit is negligible. It is thus reasonable to also assume that market orders are only sent to $Q_{\pm 1}$ and $Q_{\pm 2}$. This enables us to keep the independence property between $Q_{\pm 3}$ and $(Q_{\pm 1}, Q_{\pm 2})$. When $q_{\pm 1} > 0$, the market order intensity $\lambda^M_{buy/sell}$ is assumed to be a function of $q_{\pm 1}$; when $q_{\pm 1} = 0$, it is a function of $q_{\pm 2}$ only.

2.4.2 Model II^a: Empirical study

In this empirical study, our goal is to understand how market participants make trading decisions at $Q_{\pm 2}$ in two different situations: $q_{\pm 1} = 0$ and $q_{\pm 1} > 0$. Since we are now studying a two-dimensional problem, the data recording process is slightly different. In particular, for (Q_1, Q_2), it goes as follows: we record the waiting times $\Delta t_i(\omega)$ between events that happen at Q_1 or Q_2, the type of event $\mathcal{T}(\omega)$ and the two queue sizes $(q_1(\omega), q_2(\omega))$ before the event. The maximum

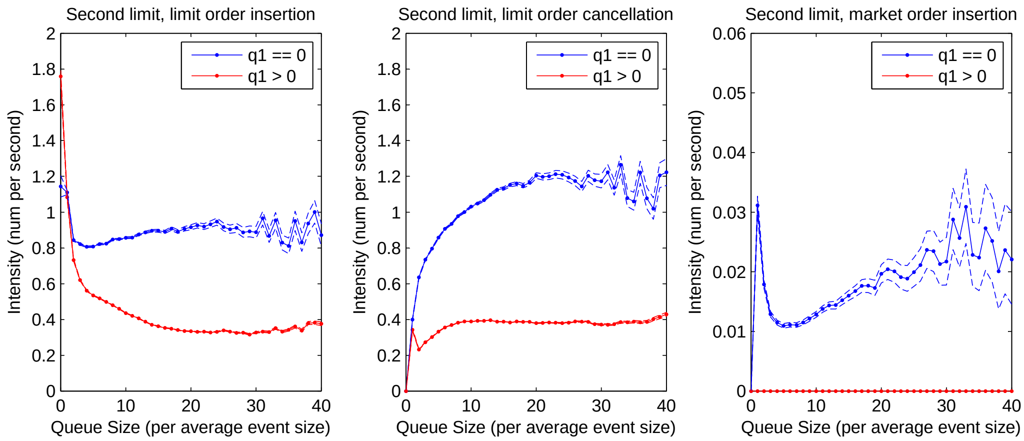

Figure I.4: Intensities at Q_2 as functions of $1_{q_1>0}$ and q_2, France Telecom

likelihood method is again used to estimate the intensity functions λ_i^L, λ_i^C, λ_i^M for $i = 1, 2$. For $i = 1$ and $i = 3$, as the dynamics at $Q_{\pm i}$ only depend on the queue size at $Q_{\pm i}$, the estimated values of λ_1^L, λ_1^C and λ_1^M are very close to those obtained in Model I and are not shown here. The estimated intensity functions at $Q_{\pm 2}$ are given in Figure I.4. Some comments are in order:

- Limit order insertion: Both curves are decreasing functions of the queue size. In the first case ($q_{\pm 1} = 0$), the limit order insertion intensity reaches very rapidly its asymptotic value. The relatively high value observed for $q_{\pm 2} = 0$ is probably due to the fact that for large tick assets, market makers rarely allow for spreads larger than 3 ticks. In the second case ($q_{\pm 1} > 0$), the intensity continues to go down to a much lower value. This is likely to be related to the arbitrage strategy introduced in Section 2.3.2: post passive orders at a non-best limit when its size is small, wait for this limit to eventually become the best limit and then gain the profit from having the priority value. For example, when the considered limit becomes the best one, one can decide to stay in the queue if its size is large enough to cover the risk of short term market trend, or to cancel the orders if the queue size is too small.

- Limit order cancellation: The cancellation rate is higher when $q_{\pm 1} = 0$. This can be related to the concentration of the trading activity at best limits. When $q_{\pm 1} > 0$, the cancellation rate is quite large when $q_{\pm 2} = 1$, as it is the case at $Q_{\pm 3}$ (see Section 2.3.2).

- Market orders: No market order can arrive at $Q_{\pm 2}$ when there are still limit orders at $Q_{\pm 1}$ (cross limits large market orders that consume several limits are treated as several market orders that arrive sequentially at those limits within a very short time period). The market order arrival rate when $Q_{\pm 2}$ is the best limit is not very different from that at $Q_{\pm 1}$, but shows a rather unexpected increasing trend when the queue size becomes larger than 5 AES_2.

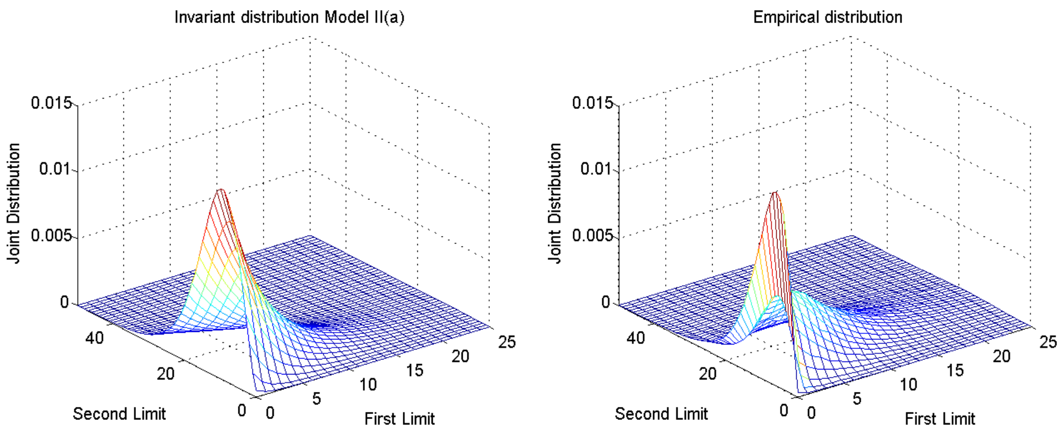

Figure I.5: Model II^a: joint distribution of q_1, q_2, France Telecom

2.4.3 Model II^a: Asymptotic behavior

Model II^a belongs to a special class of Markov processes, called quasi birth and death processes (QBD). Their asymptotic behavior can be studied by the matrix geometric method. Definitions of QBD processes and explanations about the matrix geometric method can be found in appendix. In Figure I.5, we show the theoretical joint distribution of (q_1, q_2) for the stock France Telecom and compare it with the joint distribution estimated from empirical data. Here also, we see that the theoretical results provide a very satisfying approximation.

2.4.4 Model II^b: Modeling bid-ask dependences

We now study the interactions between the bid queues and the ask queues. Let $\mathcal{Q}^0$, $\mathcal{Q}^-$, $\bar{\mathcal{Q}}$, $\mathcal{Q}^+$ be four marks which represent in the following ranges of values for the queue sizes. Let m and l be two integers. We define the function $\mathcal{S}_{m,l}(x)$:

$$\begin{aligned}
\mathcal{S}_{m,l}(x) &= \mathcal{Q}^0 \text{ if } x = 0 \\
\mathcal{S}_{m,l}(x) &= \mathcal{Q}^- \text{ if } 0 < x \leq m \\
\mathcal{S}_{m,l}(x) &= \bar{\mathcal{Q}} \text{ if } m < x \leq l \\
\mathcal{S}_{m,l}(x) &= \mathcal{Q}^+ \text{ if } x > l.
\end{aligned}$$

This function associates to a queue size x four possible ranges: empty: $x = 0$, small: $x \in (0, m]$, usual: $x \in (m, l]$ and large: $x \in (l, +\infty)$. We set m as the 33% lower quantile and l as the 33% upper quantile of $q_{\pm 1}$ (conditional on positive values). In this model, market participants at $Q_{\pm 1}$ adjust their behavior not only according to the target queue size, but also to the size of the opposite queue. The rates $\lambda^L_{\pm 1}$ and $\lambda^C_{\pm 1}$ are therefore modeled as functions of $q_{\pm 1}$ and $\mathcal{S}_{m,l}(q_{\mp 1})$. As in Model II^a, we suppose that market orders consume volume at the best limits and are only sent to $Q_{\pm 1}$ and $Q_{\pm 2}$. When $q_{\pm 1} > 0$, the market order intensity $\lambda^M_{buy/sell}$ is assumed to be a function of $q_{\pm 1}$ and $\mathcal{S}_{m,l}(q_{\mp 1})$. Regime switching at $Q_{\pm 2}$ is kept in this model: $\lambda^L_{\pm 2}$, $\lambda^C_{\pm 2}$ are assumed to be functions of $\mathbf{1}_{q_{\pm 1} > 0}$ and $q_{\pm 2}$, and when $q_{\pm 1} = 0$, the market order intensity $\lambda^M_{buy/sell}$ is modeled as a function of $q_{\pm 2}$.

Under these assumptions, the $2K$-dimensional problem is reduced to the study of the 4-dimensional continuous-time Markov jump process $(Q_{-2}, Q_{-1}, Q_1, Q_2)$. One important feature

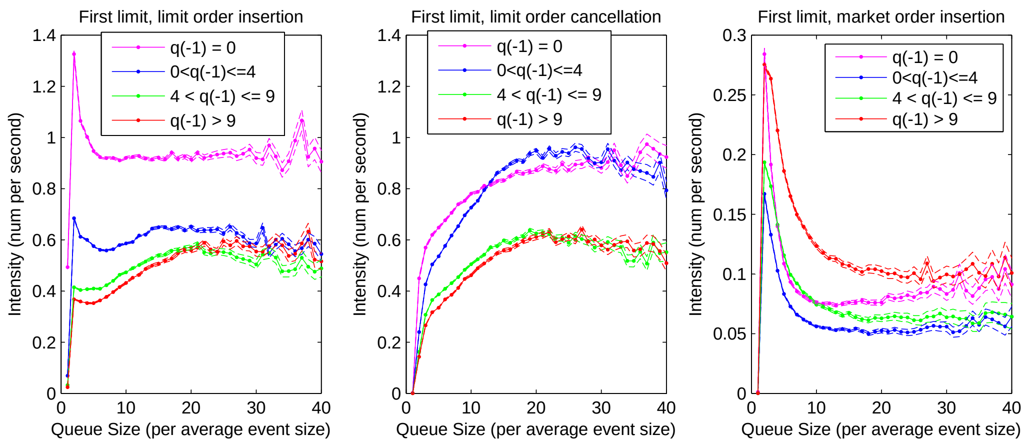

Figure I.6: Intensities at Q_1 as functions of $\mathscr{S}_{m,l}(q_{-1})$ and q_1, France Telecom

of this model is that the queues $Q_{\pm 2}$ have no influence on the dynamics at $Q_{\pm 1}$. Therefore, we only need to study the 3-dimensional process (Q_{-1}, Q_1, Q_2) (or even the 2-dimensional process (Q_{-1}, Q_1) if one is only interested in the dynamics at $Q_{\pm 1}$. Remark also that other choices for the specification of the intensity functions at $Q_{\pm 1}$ are possible. For example, one can consider them as functions of the first level bid/ask imbalance, defined as $\frac{q_1 - q_{-1}}{q_1 + q_{-1}}$, or simply as functions of the spread size.

2.4.5 Model II^b: Empirical study

We focus here on the estimation of the intensity functions at $Q_{\pm 1}$. We consider the departure flow intensities $\lambda^C_{\pm 1}(q_{\pm 1}, \mathscr{S}_{m,l}(q_{\mp 1}))$ and $\lambda^M_{buy/sell}(q_{\pm 1}, \mathscr{S}_{m,l}(q_{\mp 1}))$, and the arrival flow intensities $\lambda^L_{\pm 1}(q_{\pm 1}, \mathscr{S}_{m,l}(q_{\mp 1}))$. Using again the symmetry property of the LOB, we take $\lambda^L_1(x,y) = \lambda^L_{-1}(x,y)$, $\lambda^C_1(x,y) = \lambda^C_{-1}(x,y)$ and $\lambda^M_{sell}(x,y) = \lambda^M_{buy}(x,y)$. We record the waiting times $\Delta t(\omega)$ between events that happen at Q_1 or Q_{-1}, the types of event $\mathscr{T}(\omega)$ and the two queue sizes $(q_1(\omega), q_{-1}(\omega))$ before the event. Then we estimate these intensity functions using the maximum likelihood method. The results are shown in Figure I.6 ($m = 4$ AES$_1$, $l = 9$ AES$_1$)[8]. Some remarks are in order:

- Limit order insertion: The limit order insertion rate is a decreasing function of the opposite queue size. In particular, we see that when the opposite queue is empty (pink curve), it is significantly larger. Indeed, in that case, the “efficient" price is likely to be closer to the opposite side. Therefore limit orders at the non empty first limit are likely to be profitable.

- Limit order cancellation: The cancellation rates for different ranges of Q_{-1} are similar in their forms but have different asymptotic values. This rate is not surprisingly a decreasing function of the liquidity level on the opposite side. Indeed, when this level becomes low, many market participants cancel their limit orders and send market orders since the market is likely to move in an unfavorable direction.

- Market orders: We see that when the liquidity available on the opposite side is abundant, more market orders are sent. Indeed, in that case, transactions at the target queue are

[8]Note that the computation of the confidence intervals becomes more intricate for this model and the results presented are slightly approximate ones, see details in appendix.

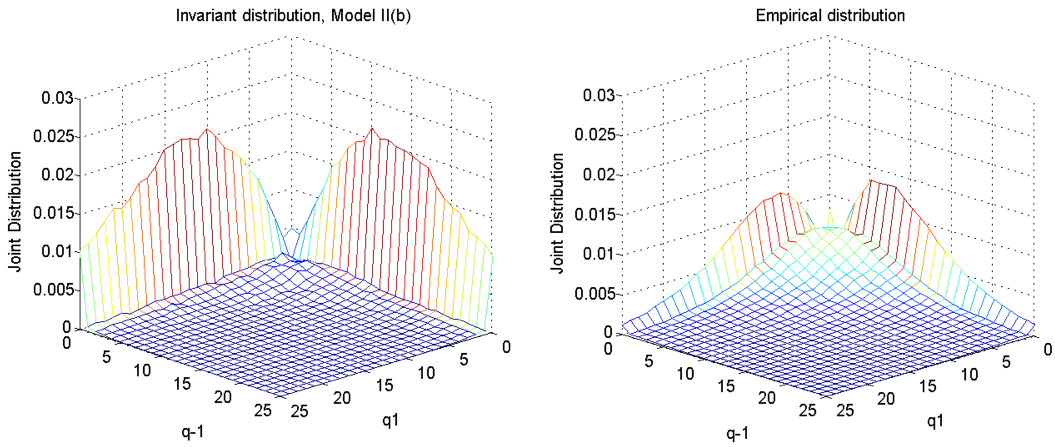

Figure I.7: Model II^b: joint distribution of q_{-1}, q_1, France Telecom

relatively cheap as its price level is temporarily closer to the efficient price. In the special situation $q_{-1} = 0$, the price level at Q_1 can seem relatively attractive since it is much closer to the reference price than the opposite best price, which is in that case 2 ticks away from it. This explains why the market order intensity is larger when the opposite queue is empty than when its size is small.

2.4.6 Model II^b: Asymptotic behavior

Monte-Carlo simulations are used to obtain the theoretical invariant distribution of the LOB in Model II^b. The theoretical and empirical joint distributions of Q_{-1} and Q_1 are shown in Figure I.7. The difference between the two graphs comes from the relatively high probabilities of states of the form (x, y) with x and y both small in empirical data, which are somehow replaced by states of the form $(x, 0)$ or $(0, y)$ in the model. Indeed, in practice, a situation where one of the first queue is empty is not likely to remain long since it often leads to a reference price change. This effect is not taken into account in Model II^b where the reference price is constant, but will be investigated in Model III in Section 3.1. We anticipate here by giving in Figure I.8 the joint distribution obtained when suitable moves of the reference price are added within the framework of Model II^b (following the approach of Model III in Section 3.1). We now find that the simulated density becomes very close to the empirical one.

2.5 Example of application: Probability of execution

The preceding models can be used to compute short term predictions about several important LOB related quantities. One relevant example is the probability of executing an order before the midprice moves. Suppose that at time $t = 0$, both Q_1 and Q_{-1} are not empty. Then a trader (called A) submits a buy limit order at Q_{-1} of size n_0 and waits in the queue until either the order is executed or the opposite queue Q_1 is totally depleted. The probability of execution can be computed in all of the three preceding models, using Monte-Carlo simulations.

There are two types of orders at Q_{-1}: orders placed before $t = 0$, thus having higher priority compared with the order of trader A, and orders placed after $t = 0$, having lower priority. When a market order arrives at Q_{-1}, the limit order with the highest priority is executed. Hence trader A's order starts being executed only when all orders placed at Q_{-1} before $t = 0$ have been

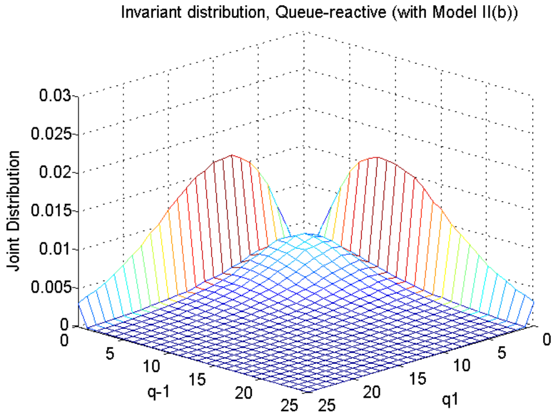

Figure I.8: Model III: joint distribution of q_{-1}, q_1, France Telecom

either canceled or executed. When a cancellation event happens at Q_{-1}, the precise order being canceled is not clearly defined in our models. So, we need to make two additional assumptions for the cancellation process.

Assumption 18. *When a cancellation event occurs at Q_{-1}, orders at Q_{-1} have the same probability of being canceled (except for the limit order submitted by trader A, which is never canceled).*

Assumption 19. *The cancellation intensity at Q_{-1} is supposed to be equal to $\lambda_1^C(q_{-1})\frac{q_{-1}-n_0}{q_{-1}}$ instead of $\lambda_1^C(q_{-1})$, since the order placed by trader A is never canceled.*

Orders with lower priority are actually more likely to be canceled, see Gareche, Disdier, Kockelkoren, and Bouchaud (2013). However, in order to investigate precisely this feature, we would need more detailed market data keeping records of the identifiers of the submitted and canceled orders. As a result, execution probabilities might be slightly overestimated using Assumptions 18 and 19. Simulation results (for $n_0 = 1$) are shown in Figure I.9, together with the predictions associated to a Poisson model that assumes a linearly increasing cancellation rate. We see that our three models give fairly similar execution probabilities, while the Poisson model clearly overestimates them.

3 The queue-reactive model: a time consistent model with stochastic LOB and dynamic reference price

We now wish to obtain a model which is relevant on the whole period of interest and provides useful applications.

3.1 Model III: The queue-reactive model

3.1.1 Building the model

Let δ denote the tick value. We assume here that p_{ref} changes with some probability θ when some event modifies the midprice p_{mid}. More precisely, when p_{mid} increases/decreases[9], p_{ref}

[9]Note that in this model, p_{ref} does not necessarily match its estimated value using the method introduced in Section 2.2. However, for large tick assets, the difference is negligible.

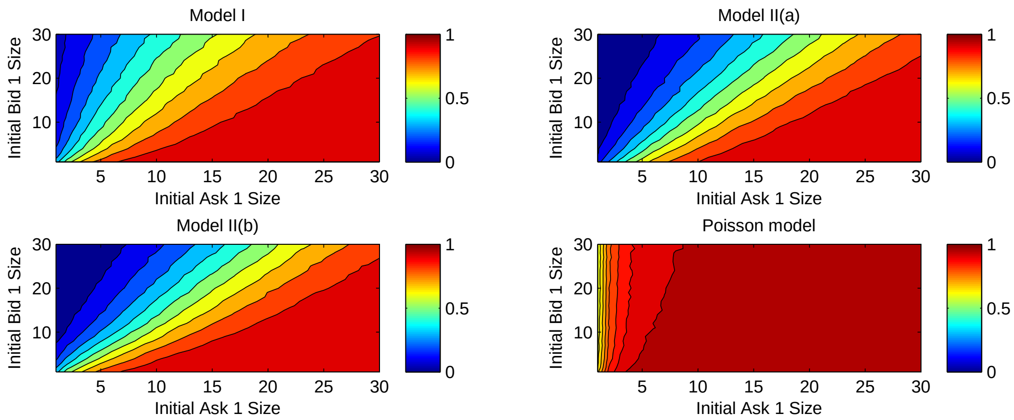

Figure I.9: Execution probability of a buying order placed at Q_{-1} at $t = 0$, France Telecom

increases/decreases by δ with probability θ, provided $q_{\pm 1} = 0$ at that moment. Hence changes of p_{ref} are possibly triggered by one of the three following events:

- The insertion of a buy limit order within the bid-ask spread while Q_1 is empty at the moment of this insertion, or the insertion of a sell limit order within the bid-ask spread while Q_{-1} is empty at the moment of this insertion.
- A cancellation of the last limit order at one of the best offer queues.
- A market order that consumes the last limit order at one of the best offer queues.

When p_{ref} changes, the value of q_i switches immediately to the value of one of its neighbors (right if p_{ref} increases, left if it decreases). Thus, $q_{\pm 1}$ becomes zero when p_{ref} decreases/increases. Recall that we keep records of the LOB up to the third limit. Consequently, the value for $q_{\pm 3}$ when p_{ref} increases/decreases is drawn from its invariant measure. Note that the queue switching process must be handled very carefully: the average event sizes are not the same for different queues. So, when q_i becomes q_j, its new value should be re-normalized by the ratio between the two average event sizes at Q_i and Q_j.

To possibly incorporate external information, we moreover assume that with probability θ^{reinit}, the LOB state is redrawn from its invariant distribution around the new reference price when p_{ref} changes. The parameter θ^{reinit} can be understood as the percentage of price changes due to exogenous information. In this case, we consider that market participants readjust very quickly their order flows around the new reference price, as if a new state of the LOB was drawn from its invariant distribution. A similar approach has been used in Cont and De Larrard (2013) in a model for best bid and best ask queues, in which θ^{reinit} is set to 1. Under these assumptions, the market dynamics is now modeled by a $(2K+1)$-dimensional Markov process: $\tilde{X}(t) := (X(t), p_{ref}(t))$, in the countable state space $\tilde{\Omega} = \mathbb{N}^{2K} \times \delta\mathbb{N}$, where $X(t) = (q_{-K}(t), \ldots, q_{-1}(t), q_1(t), \ldots, q_K(t))$ represents the available volumes at different limits.

In the sequel, Model I is used to describe the LOB dynamics during periods when p_{ref} is constant (very similar results are obtained in simulations using Model II^a or II^b). The p_{ref} change probability θ and the LOB reinitialization probability θ^{reinit} are calibrated using the 10

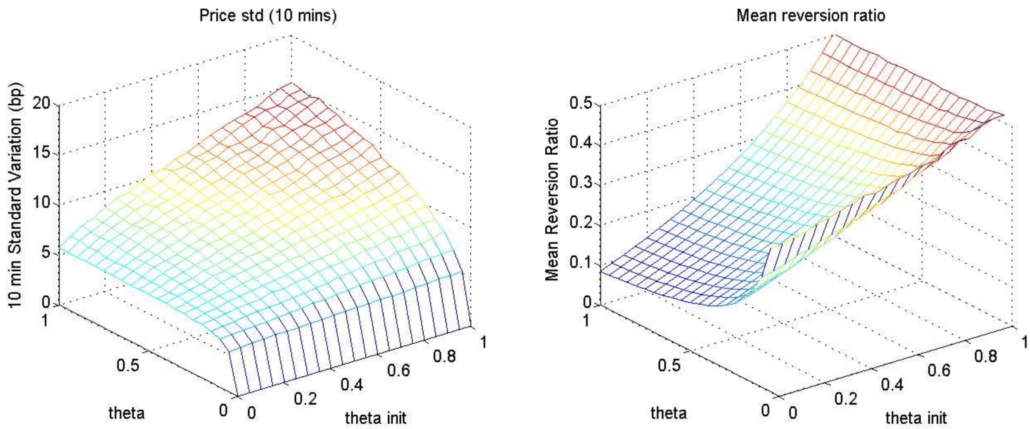

Figure I.10: 10 min volatility and mean reversion ratio, France Telecom

minutes standard deviation of the returns of p_{mid} (the volatility) and the mean reversion ratio η introduced in Robert and Rosenbaum (2011), defined by

$$\eta = \frac{N_c}{2N_a},$$

where N_c is the number of continuations of the estimated p_{ref} on the interval of interest (that is the number of consecutive moves in the same direction) and N_a is the number of alternations (that is the number of consecutive moves in opposite directions)[10]. Indeed, the microstructure of large tick assets is well summarized by the parameter η, see Robert and Rosenbaum (2011) and Dayri and Rosenbaum (2012) and the volatility is of course one of the most important low frequency statistics. In Figure I.10, we show the surfaces of the 10 min volatility and η for different values of θ and θ^{reinit}.

3.1.2 About the maximal mechanical volatility

Let us comment now the particular case where we take $\theta^{reinit} = 0$. In such situation, Model III becomes a “purely order book driven model" since the price fluctuations are completely generated by the LOB dynamics. Our simulations show that under this setting, the maximal attainable volatility level (when $\theta = 1$), which we call maximal mechanical volatility, is much lower than the empirical volatility (5 bps compared with 14 bps for the stock France Telecom). This suggests that endogenous LOB dynamics alone may not be enough for reproducing the market volatility. A closer look at these results shows that the model approximates actually quite well the average frequency of price changes, and that the small value of the mechanical volatility is mainly due to the strong mean reverting behavior of the price in this purely order book driven model. This is because of the often reversed bid-ask imbalance immediately after a change of p_{ref}. In Figure I.10, we can see that the mean reversion ratio η is equal to 0.08 when $\theta = 1, \theta^{reinit} = 0$, which is much smaller than the empirical ratio 0.39.

[10] Note that here we compute the mean reversion ratio of p_{ref} while the transaction price is usually considered.

3.2 Example of application: Order placement analysis

We now show how the queue-reactive model can be used in the context of optimal trading. In the general framework of optimal execution, the trading horizon is split into small slices (typically 5-10 minutes) and an algorithm of execution determines the volume to be executed in each slice. This problem, often called "order scheduling problem", has been widely studied in the literature, see Bertsimas, Lo, and Hummel (1999); Almgren and Chriss (2000); Bouchard, Dang, and Lehalle (2011) for representative examples on this topic. In practice, another optimization issue, the "order placement problem", arises naturally once the order scheduling problem has been solved: how should the algorithm place orders to execute the target volume?

This second optimization problem can be seen as the microstructural version of the first one. However, it is much more difficult to solve. Indeed, the price dynamics can no longer be approximated by a Brownian motion at these (ultra) high frequency scales. Moreover, the queue priority plays an important role as well as other microstructural features of the asset such as the tick size, the state of the LOB and the trading speed. This in particular implies that execution strategies based on limit or market orders can lead to very different outcomes. Some papers investigate the consequences of using different types of order, see Harris and Hasbrouck (1996), while others aim at finding the best position to place limit orders, see Laruelle, Lehalle, and Pagès (2013). However, in practice, order placement tactics are usually more complex than the ones considered in the academic literature. For example, traders can hide their trading intentions by splitting furthermore the target volume within each slice. Also, they may start passively, sending limit orders, and then switch to market orders when some market conditions change or a stopping time criteria is met. Very few quantitative tools are available for the analysis of sophisticated tactics and one often needs to rely on so-called market replayers, in which the number of simulations is limited by that of the available trading days in the historical data. Moreover, the market impact, that is the average price drift due to our own trading between the beginning of the execution and a later time, is often neglected. In contrast, our framework is unlimited in number of simulations and is both relevant and easy to use in order to study market impact profiles and execution costs of complex placement tactics.

We write n_{total} for the total quantity to execute and M for the number of slices. An order scheduling strategy gives the target quantity to be executed in each slice, denoted by n_i ($n_i \geq 0$ and $\sum_{i=1}^{M} n_i = n_{total}$). An order placement tactic can be seen as a predefined procedure of order management, ensuring the execution of the target quantity within the slice. Here, as illustration examples, we present two simple tactics, denoted by **T**1 and **T**2. In the i-th slice, both tactics post a limit order of size n_i at the best offer queue at the beginning of the period, and send a market order with all the remaining quantity to complete the execution of the target volume at the end time of the slice. In between:

- **T**1 (Fire and forget): When p_{mid} changes, cancel the limit order and send a market order at the opposite side with all the remaining volume if any.
- **T**2 (Pegging to the best): When the best offer price changes or our order is the only remaining order at the best offer limit, cancel the order and repost all the remaining volume at the newly revealed best offer queue.

Since an order placement tactic is often specifically designed for a given order scheduling strategy, comparisons between two tactics should take into account the associated scheduling strategy, together with the target benchmark[11]. Other parameters can also have influence when comparing two tactics, such as the total quantity to execute n_{total} and the number of slices M. To simplify our analysis, we simulate a buy order of size $n_{total} = 60$ AES$_1$, with $M = 20$ and the duration of each slice is fixed to 10 minutes (a total trading period of 3h20). We focus on two benchmarks: the VWAP on the total period (volume weighted average transaction price) and the arrival price S_0 (the midprice when the execution algorithm starts). Moreover, two types of order scheduling strategies, denoted by **S**1 and **S**2, are considered to partly reflect the diversity of optimal trading schemes:

- **S**1: A linear scheduling ($n_i = n_{total}/M$), used for the VWAP benchmark.
- **S**2: An exponential scheduling $n_i = n_{total}(e^{-(i-1)/4} - e^{-i/4})$, used for the benchmark S_0.

Finally, note that Assumptions 18 and 19 are in force for the order cancellation processes.

3.2.1 Tactic performance analysis

The performance of an execution algorithm is often measured by its slippage, defined (for a buy order) by

$$\text{Slippage} = \frac{P_{benchmark} - P_{exec}}{P_{benchmark}}.$$

To understand the effects of the order placement tactic on the execution's slippage, we define the theoretical scheduling slippage by:

$$\begin{aligned} P_{exec}^{theo} &= \sum_{i=1}^{M} n_i \text{VWAP}^i \\ \text{Slippage}^{theo} &= \frac{P_{benchmark} - P_{exec}^{theo}}{P_{benchmark}}, \end{aligned}$$

where VWAP^i denotes the volume weighted average transaction price of the $i-$th slice. Indeed, VWAP^i is often considered as a simple proxy for the execution price in the slice when one focuses on the scheduling algorithm. Hence, Slippage^{theo} essentially measures the quality of the scheduling strategy and neglects the randomness in execution prices due to the order placement tactic[12]. Note that here a market impact component is included in the computation of the theoretical scheduling slippage. This is because the value of VWAP^i in each slice is obviously impacted by our execution.

We launched 2000 simulations for each couple of (**S**1/S2, **T**1/T2). The intensity functions estimated for the stock France Telecom are used in these simulations, as well as the two parameters $\theta = 0.7$ and $\theta^{reinit} = 0.85$ calibrated in Section 3.1. Furthermore, we use a standard kernel smoothing method when estimating the probability density functions of Slippage^{theo} and Slippage. The results are shown in Figure I.11.

Figure I.11 suggests that the slippage distributions of the same scheduling strategy using two different tactics can be very different: **T**2 ("Pegging to the best") performs better than **T**1 ("Fire

[11]In execution services, the client often wants the execution algorithm to target some specific price (the arrival price, the average market price during a predefined period,...). The quality of the execution is then assessed on the basis of the difference between the realized execution price and this target benchmark price.

[12]See Section 5.4 in Appendix for a detailed discussion on placement tactic analysis

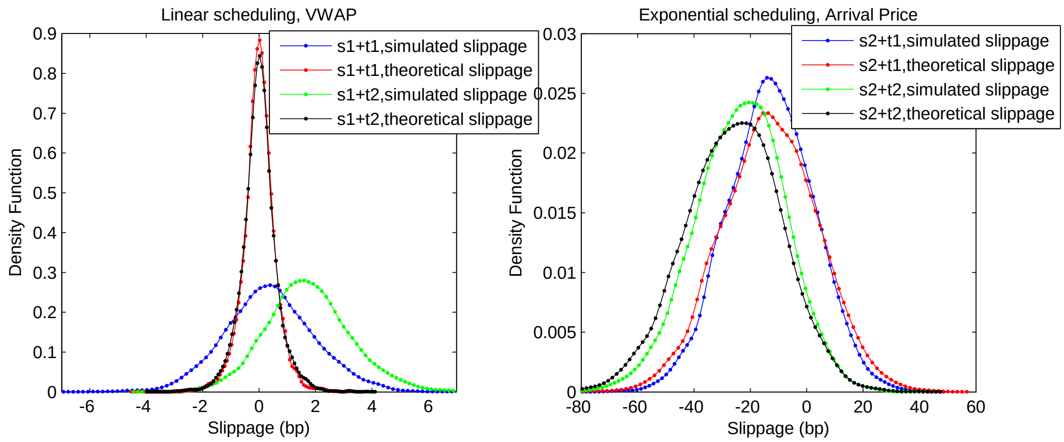

Figure I.11: Simulation results for the tactics

and forget") when being coupled with a linear scheduling strategy with VWAP benchmark, while **T**1 slightly outperforms **T**2 when an exponential scheduling strategy with arrival price benchmark is considered. In our setting, the limit orders change the queue sizes and therefore modify the behaviors of the order flows. Consequently they generate market impact. By constantly following the best offer queue until the total volume is filled, **T**2 achieves on average a higher passive execution rate (defined as the volume passively executed[13] divided by the total executed volume). Thus, in each slice, it often obtains a better price than that of a more market orders based tactic. However, at the same time, it creates a larger impact than **T**1 since the order stays longer in the queues. This explains why the theoretical scheduling slippage of **T**2 is worse than that of **T**1 for an execution with arrival price benchmark using an exponential scheduling strategy.

3.2.2 Market impact profiles

We now study the market impact profiles of these two tactics. Recall that an order placement tactic has two parameters: the slice duration T and the quantity to execute n. In the following experiments, T is set to 10 minutes, and the value of n varies from 1 to 60 AES_1. We denote by $\mathrm{MI}^i(t,n)$ the market impact at time t of Tactic i with target quantity n, defined by: $\mathrm{MI}^i(t,n) = \mathrm{E}[\frac{S_t - S_0}{S_0}]$, with S_t the midprice at time t. We launched 2000 simulations for each value of n, t in the ranges 1-60 AES_1 and 1-600 seconds. Impact profiles are given in Figure I.12.

In agreement with the celebrated "square-root law", see Gatheral (2010); Toth, Lemperiere, Deremble, De Lataillade, Kockelkoren, and Bouchaud (2011a); Farmer, Gerig, Lillo, and Waelbroeck (2013), the market impact curves are concave both in time and volume. One can also see that the impact of **T**1 is quite instantaneous and depends essentially on the target quantity n, while the impact of **T**2 is a progressive process, depending both on the target quantity n and the time t. Remark that **T**2 seems suitable when dealing with small orders since its market impact is small and it has a higher passive execution rate than **T**1. If one needs to trade larger orders, **T**1 becomes probably more relevant since the cost of market impact is likely to outweigh the benefit from passive execution of **T**2. Finally, note that in our Markovian framework, no significant price relaxation (that is the fact that on average, after the completion of the execution of a buy

[13] A buy execution is said to be passive if it occurs at the bid side of the LOB, aggressive if it occurs at the ask side of the LOB.

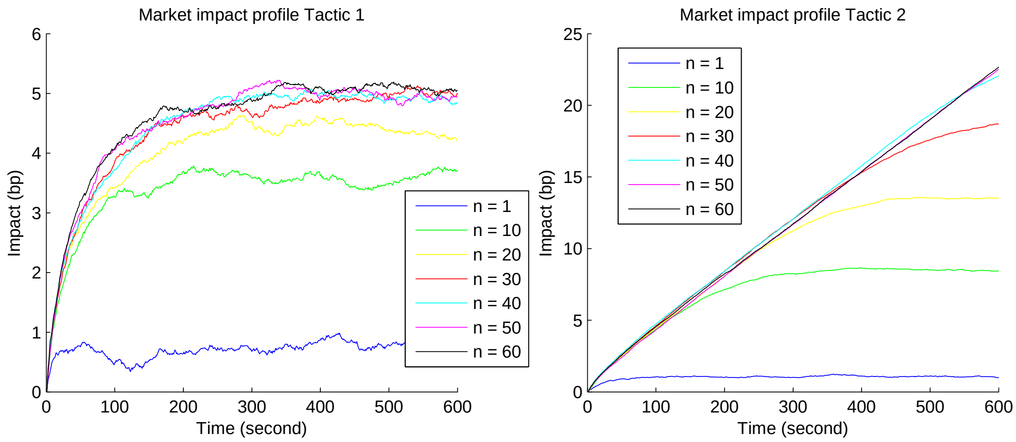

Figure I.12: Market impact profiles

order, the price may drop to a lower level than the one reached at the end of the execution) can be observed.

4 Conclusion and perspectives

In this work, we have modeled market participants intelligence through their average behaviors towards various states of the LOB. This enabled us to analyze the different order flows and to design a suitable market simulator for practitioners, allowing notably to investigate the transaction costs of complex trading strategies. To our knowledge, our model is the first one where such pre-trade cost analysis is possible in a simple and efficient way.

Another important public information, the historical order flow, is not considered in this approach. Market order flows have been shown to be autocorrelated in several empirical studies, see for example Toth, Palit, Lillo, and Farmer (2011b). Thus, adding such feature in our framework would probably be relevant. Another possible direction for future research would be to explain the shape of the estimated intensity functions in a more sophisticated way. For example, it would be interesting to design some agent based model where these repetitive patterns of the LOB dynamics would be reproduced, providing an even better understanding of the nature of these intensity curves.

5 Appendix

5.1 Proof of Theorem 1

Proof. For some $z > 1$, set

$$V(q) = \sum_{j=-K, j\neq 0}^{K} z^{|q_j - C_{bound}|_+}.$$

For any $q \in \Omega$, we have:

$$\begin{aligned}
\mathcal{Q}V(q) &= \sum_{p\neq q} \mathcal{Q}_{q,p}[V(p) - V(q)] \\
&= \sum_{i=-K,i\neq 0}^{K} [f_i(q)(z^{|q_i+1-C_{bound}|_+} - z^{|q_i-C_{bound}|_+}) + g_i(q)(z^{|q_i-1-C_{bound}|_+} - z^{|q_i-C_{bound}|_+})] \\
&= \sum_{i=-K,i\neq 0}^{K} [f_i(q)1_{q_i\geq C_{bound}} z^{|q_i-C_{bound}|_+}(z-1) + g_i(q)1_{q_i\geq C_{bound}+1} z^{|q_i-C_{bound}|_+}(\frac{1}{z}-1)] \\
&= (z-1)\sum_{i=-K,i\neq 0}^{K} [f_i(q)1_{q_i\geq C_{bound}} - \frac{g_i(q)1_{q_i\geq C_{bound}+1}}{z}] z^{|q_i-C_{bound}|_+} \\
&= (z-1)\sum_{i:q_i=C_{bound}} f_i(q) + (z-1)\sum_{i:q_i>C_{bound}} [f_i(q) - \frac{g_i(q)}{z}] z^{q_i-C_{bound}}. \qquad \text{(I.1)}
\end{aligned}$$

Under Assumption 16 and 17, we can find a z sufficiently close to 1 such that, if $q_i > C_{bound}$,

$$f_i(q) - \frac{g_i(q)}{z} < z^{-1}(-r + H(z-1)) = -r' < 0.$$

So, from Equation (I.1), we have

$$\begin{aligned}
\mathcal{Q}V(q) &\leq (z-1)H - (z-1)\sum_{i:q_i>C_{bound}} r' z^{q_i-C_{bound}} \\
&\leq -(z-1)r'\sum_i z^{|q_i-C_{bound}|_+} + (z-1)H + 2(z-1)r'K \\
&\leq -(z-1)r'V(q) + (z-1)[H + 2r'Kz].
\end{aligned}$$

Thus $X(t)$ is V-uniformly ergodic. Then using Theorem 4.2 in Meyn and Tweedie (1993), $X(t)$ is Harris positive recurrent and has a finite invariant measure. Furthermore, by Theorem 3.6.2 in Norris (1998), the process $X(t)$ converges to its equilibrium and is therefore ergodic. □

5.2 Computation of confidence intervals

When the queues are independent, by the central limit theorem, we have, with asymptotic probability 95% (we note $\hat{p}_i{}^L(n) = \frac{\#\{\mathcal{T}(\omega)\in\mathcal{E}^+, q_i(\omega)=n\}}{\#\{q_i(\omega)=n\}}$):

$$\begin{aligned}
\Lambda_i(n) &\in [\hat{\Lambda}_i(n) - \frac{1.96\hat{\Lambda}_i(n)}{\sqrt{\#\{q_i(\omega)=n\}}}, \hat{\Lambda}_i(n) + \frac{1.96\hat{\Lambda}_i(n)}{\sqrt{\#\{q_i(\omega)=n\}}}] \\
\frac{\lambda_i^L(n)}{\Lambda_i(n)} &\in [\hat{p}_i{}^L(n) - \frac{1.96\sqrt{\hat{p}_i{}^L(n)(1-\hat{p}_i{}^L(n))}}{\sqrt{\#\{q_i(\omega)=n\}}}, \hat{p}_i{}^L(n) + \frac{1.96\sqrt{\hat{p}_i{}^L(n)(1-\hat{p}_i{}^L(n))}}{\sqrt{\#\{q_i(\omega)=n\}}}].
\end{aligned}$$

So, at least with probability 90%:

$$\begin{aligned}\lambda_i^L(n) \quad \in \quad & [(\hat{\Lambda}_i(n) - \frac{1.96\hat{\Lambda}_i(n)}{\sqrt{\#\{q_i(\omega)=n\}}})(\hat{p}_i{}^L(n) - \frac{1.96\sqrt{\hat{p}_i{}^L(n)(1-\hat{p}_i{}^L(n))}}{\sqrt{\#\{q_i(\omega)=n\}}}) \\ & (\hat{\Lambda}_i(n) + \frac{1.96\hat{\Lambda}_i(n)}{\sqrt{\#\{q_i(\omega)=n\}}})(\hat{p}_i{}^L(n) + \frac{1.96\sqrt{\hat{p}_i{}^L(n)(1-\hat{p}_i{}^L(n))}}{\sqrt{\#\{q_i(\omega)=n\}}})].\end{aligned}$$

Similar results can be computed for λ_i^C and λ_i^M. The method used to compute confidence intervals of Model II^a is quite similar. Confidence intervals are more difficult to compute in Model II^b, and we use approximations by neglecting the possible intersections between the two sets: $\{q_1(\omega)=n, \mathscr{S}_{m,l}(q_{-1}(\omega)) \in s)\}$ and $\{q_{-1}(\omega)=n, \mathscr{S}_{m,l}(q_1(\omega)) \in s)\}$.

5.3 Quasi birth and death process

Definition 1. *(Quasi birth and death process, from Latouche and Ramaswami (1999)): A quasi birth and death (QBD) process is a bivariate Markov process with countable state space $S = \{(i,j) : i \geq 0, j = 0,1,...,m\}$ where the first element i is called the level of the process, and the second element j is called the phase of the process. The parameter m can be either finite or infinite. The process is restricted in level jumps only to its nearest neighbors, meaning that the probability of jumping from level i directly to level $l, l \geq i+2$ or $l \leq i-2$ is equal to zero.*

We can easily see that the Markov process (q_1, q_2) in Model II^a is indeed a QBD process with countable phases. Its infinitesimal generator matrix is of the following form:

$$Q = \begin{bmatrix} A_1^{(0)} & A_0^{(0)} & 0 & 0 & \dots \\ A_2^{(1)} & A_1^{(1)} & A_0^{(1)} & 0 & \dots \\ 0 & A_2^{(2)} & A_1^{(2)} & A_0^{(2)} & \dots \\ \dots & \dots & \dots & \dots & \dots \end{bmatrix},$$

where the matrix $A_0^{(\ell)}$ encodes transitions from level $q_1 = \ell$ to level $q_1 = \ell+1$, matrix $A_2^{(\ell)}$ encodes transitions from level $q_1 = \ell$ to level $q_1 = \ell - 1$, and matrix $A_1^{(\ell)}$ encodes transitions within level $q_1 = \ell$. More specifically, the element (i,j) of $A_0^{(\ell)}$ is the transition rate from state $(q_1 = \ell, q_2 = i)$ to state $(q_1 = \ell+1, q_2 = j)$, the element (i,j) of $A_2^{(\ell)}$ is the transition rate from state $(q_1 = \ell, q_2 = i)$ to state $(q_1 = \ell-1, q_2 = j)$, and the element (i,j) of $A_1^{(\ell)}$ is the transition rate from state $(q_1 = \ell, q_2 = i)$ to state $(q_1 = \ell, q_2 = j)$.

We write the intensity functions at Q_2 when $q_1 = 0$ with a $\tilde{}$. For matrix $A_i^{(\ell)}$, $i = 0,1,2$, we have:

$$\begin{aligned} A_0^{(k)} &= \lambda_1^L(k) I, \\ A_2^{(k)} &= (\lambda_1^C(k) + \lambda_{buy}^M(k)) I, \end{aligned}$$

$$A_1^{(0)} = \begin{bmatrix} -\lambda_1^L(0) - \tilde{\lambda}_2^L(0) & \tilde{\lambda}_2^L(0) & 0 & \dots \\ \tilde{\lambda}_2^C(1) + \tilde{\lambda}_{buy}^M(1) & -\lambda_1^L(0) - \tilde{\lambda}_2^L(1) - \tilde{\lambda}_2^C(1) - \tilde{\lambda}_{buy}^M(1) & \tilde{\lambda}_2^L(1) & \dots \\ 0 & \tilde{\lambda}_2^C(2) + \tilde{\lambda}_{buy}^M(2) & -\lambda_1^L(0) - \tilde{\lambda}_2^L(2) - \tilde{\lambda}_2^C(2) - \tilde{\lambda}_{buy}^M(2) & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix},$$

and for $k \geq 1$:

$$A_1^{(k)} = \begin{bmatrix} -\lambda_1^C(k) - \lambda_{buy}^M(k) - \lambda_1^L(k) - \lambda_2^L(0) & \lambda_2^L(0) & 0 & \dots \\ \lambda_2^C(1) & -\lambda_1^C(k) - \lambda_{buy}^M(k) - \lambda_1^L(k) - \lambda_2^L(1) - \lambda_2^C(1) & \lambda_2^L(1) & \dots \\ \dots & \dots & \dots & \dots \end{bmatrix}.$$

We define $\pi_{i,j} = \mathbb{P}[q_1 = i, q_2 = j]$ the stationary distribution of this QBD process, and:

$$\begin{aligned}\pi_n &= [\pi_{n,0}, \pi_{n,1}, ...]\\ \pi &= [\pi_0, \pi_1, ...].\end{aligned}$$

We shall have:

$$\begin{aligned}\pi Q &= 0\\ \pi 1 &= 1.\end{aligned}$$

The dynamics of the two queues system (q_1, q_2) is level dependent, meaning that its transition kernel depends on the value of q_1. This makes the computation or approximation of its asymptotic behavior quite difficult. Thus we consider an additional assumption in order to turn (q_1, q_2) into a so-called level independent QBD process. This is particularly interesting since it enables us to easily express the invariant measure in a matrix geometric form and to compute it numerically. The level independence property is defined by the fact that for $i \geq 1$, $A_0^{(i)}$, $A_1^{(i)}$ and $A_2^{(i)}$ do not depend on i and $A_0^{(0)} = A_0^{(i)}$, see Latouche and Ramaswami (1999). Under the following assumption, this property is satisfied by (q_1, q_2) in Model II^a.

Assumption 20. *(Independent Poisson flows at first limits) There are two positive constants λ_1 and μ_1, with $\lambda_1 < \mu_1$, such that for $k \geq 1$:*

$$\begin{aligned}\lambda_1^C(k) + \lambda_{buy}^M(k) &= \mu_1\\ \lambda_1^L(k) &= \lambda_1\\ \lambda_1^L(0) &= \lambda_1.\end{aligned}$$

In practice, λ_1 and μ_1 are taken as the average values of the estimated intensity functions at first limits. Under this assumption, a quite simple numerical computation of the invariant distribution is possible. Generally speaking, QBD processes with finite phase (meaning that the value set of the second dimension, in our case q_2, is finite) can be easily treated, see for example Latouche and Ramaswami (1999). In the infinite case, truncation methods must be applied to obtain approximate results. Thanks to the special structure of the generator in our model, one simple truncation method, called "first column augmentation by block", can be applied. Details of this truncation method can be found in Bean and Latouche (2010). The mathlab toolbox SMCSolver, see Bini, Meini, Steffé, and Van Houdt (2006), is used to compute the invariant measure.

5.4 Order Placement Tactic Analysis

In this section we gather our ideas on order placement tactic analysis related to Section 3.2. We use the following notations:

M Number of slices.

n_{total} Total target execution quantity.

V Total market volume in the $i-$th slice.

n_i Target volume in the $i-$th slice.

v_i Market volume in the $i-$th slice.

n_i^* Executed volume in the $i-$th slice.

Δn_i := $n_i^* - n_i$.

N_i := $\sum_{j=1}^{i} n_i$ Cumulative target volume till the $i-$th slice.

N_i^* := $\sum_{j=1}^{i} n_i^*$ Cumulative executed volume till the $i-$th slice.

ΔN_i := $N_i^* - N_i$.

p_i Market VWAP (volumed weighted average price) in the $i-$th slice.

p_i^* Algorithm's VWAP in the $i-$th slice.

Δp_i := $p_i^* - p_i$.

σ_i Price volatility in the $i-$th slice.

In practice, an execution algorithm gives answers to the two following question:

1. Order Scheduling: how to distribute the target volume across the trading horizon?
2. Order Placement: how to place individual orders to the LOB?

While most existing approaches for post-trade performance analysis focus on evaluating the overall performance, it is actually more reasonable to separate the order scheduling part from the order placement part: performance of order placement tactics depends mainly on ultra-high frequency factors such as the latency, the queue priority, bid-ask imbalance, which have generally little influences over the choice of the "optimal trading curve" determined by the order scheduling strategy. Moreover, the same order scheduling strategy can be coupled with different order placement tactics to build different execution algorithms. In such cases, it is important to be able to understand the pros and cons of each tactic so that an informed choice can be made to determine the best tactic under different market conditions.

5.4.1 Slippage Decomposition

To separate effects of order scheduling and order placement, we propose the following decomposition method (taking a VWAP benchmark ($P_{\text{benchmark}} = \sum_{i=1}^{M} \frac{v_i}{V} p_i$) as an example, decomposition methods for other benchmark are similar):

$$\begin{aligned}
\text{Slippage} &= \frac{\sum_{i=1}^{M}(\frac{n_i^*}{n_{total}} p_i^* - \frac{v_i}{V} p_i)}{P_{\text{benchmark}}} \\
&= \frac{\sum_{i=1}^{M}(\frac{n_i}{n_{total}} - \frac{v_i}{V}) p_i}{P_{\text{benchmark}}} + \frac{(\sum_{i=1}^{M} \frac{\delta n_i}{n_{total}} p_i + \sum_{i=1}^{M} \frac{n_i}{n_{total}} \delta p_i + \sum_{i=1}^{M} \frac{\delta n_i}{n_{total}} \delta p_i)}{P_{\text{benchmark}}} \\
&= \text{Slippage}^V + \text{Slippage}^O \\
\text{Slippage}^V &= \frac{\sum_{i=1}^{M}(\frac{n_i}{n_{total}} - \frac{v_i}{V}) p_i}{P_{\text{benchmark}}} \\
\text{Slippage}^O &= \frac{\sum_{i=1}^{M} \frac{\delta n_i}{n_{total}} p_i}{P_{\text{benchmark}}} + \frac{\sum_{i=1}^{M} \frac{n_i}{n_{total}} \delta p_i}{P_{\text{benchmark}}} + \frac{\sum_{i=1}^{M} \frac{\delta n_i}{n_{total}} \delta p_i}{P_{\text{benchmark}}} \\
&= \text{Tracking Error} + \text{Price Improvement} + \text{Residual}. \qquad \text{(I.2)}
\end{aligned}$$

Volume Scheduling

Slippage^V [14] represents the quality of the volume scheduling strategy, its randomness comes from the randomness of the price process p^i and the market volume curve m_i. While traditional performance analysis uses often the deviation between the optimal trading curve and the realized market volume curve to quantify the VWAP execution's performance, we see clearly that here the quality of order scheduling depends not only on the difference between the vectors $\frac{v_i}{V}$ and $\frac{m_i}{M}$, but also on the price process p_i.

Order Placement

An order placement tactic can be seen as a (dynamic or statical) strategy of placing orders to markets with some target quantity and a limited trading horizon. Mathematically, it can be represented by three random processes: $[f(t), p(t), i(t)]$, where $f(t)$ denotes the fill rate (the executed quantity) of the tactic till time t, $p(t)$ denotes the tactic's relative performance with respect to the market till time t, and $i(t)$ denotes the market impact caused by the tactic. When other execution conditions, such as the target quantity and the duration, are equal, then we can further reduce the representation form of a placement tactic to three random variables $[f, p, i]$. The decomposition formula I.2 enables us to differentiate the contributions of three main aspects in an order placement tactics: tracking error, price improvement and the residual:

- Tracking Error: Tracking error measures the performance due to the deviation of the realized execution trading curve from the scheduled one. In the order placement analysis in Section 3.2, this term is equal to zero, as both tactics execute the exact amount assigned by the scheduling strategy in each trading slice. In practice, a good algorithm does not necessarily follows strictly the “optimal trading curve”, since this often obliges the algorithm to send a lots of market orders, which is of course not the best way to obtain the liquidity. By allowing some deviations around the optimal curve, an algorithm can have a better performance (less aggressive orders + ability to auto-adjust its trading curve under different market conditions). Tracking error is closely linked with the fill rate (f) of the placement tactic.

- Price Improvement: Price improvement measures the placement tactic's ability of capturing cheap liquidities in the microscopic level. In general, placement tactics with higher aggressive ratio (defined as the percentage of quantities executed as market orders) have worse average price. Price improvement is closely linked with the relative performance p of the placement tactic.

- Residual: Residual measures the cross effects due to the dependencies between the fill rate f and the relative performance p.

5.5 Alcatel-Lucent

Results for the stock Alcatel-Lucent are presented in the following figures (Figure I.13 to Figure I.18).

5.6 AES

AES$_i$ is defined as the average size of all events (including limit order insertion, cancellation and trades) at Q_i, while ATS computes only the average size of all trade events. In Table I.2 we show

[14]One may notice that Slippage^{theo} is exactly Slippage^V in the above decomposition.

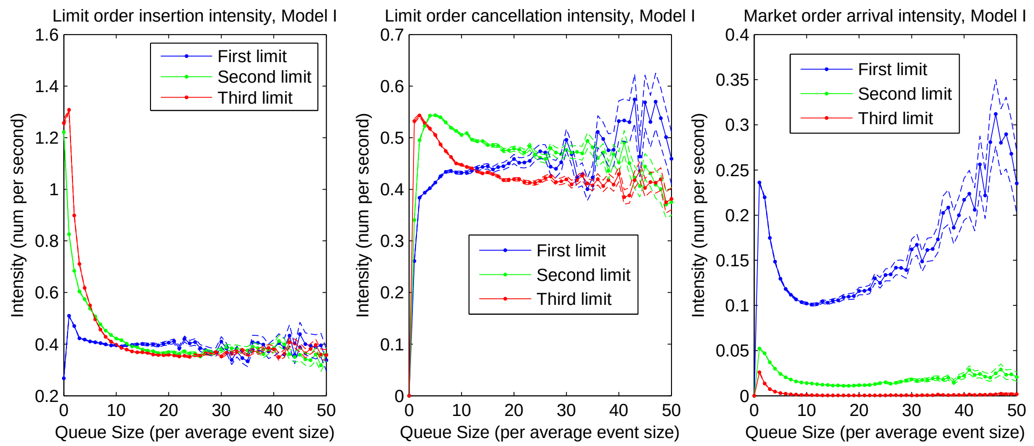

Figure I.13: Intensities at $Q_{\pm 1,2,3}$, Alcatel Lucent

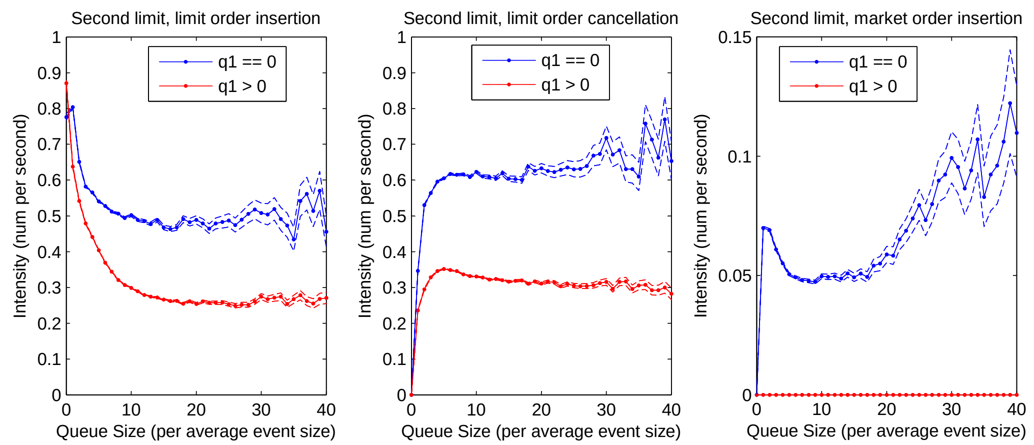

Figure I.14: Intensities at Q_2 as functions of $1_{q_1>0}$ and q_2, Alcatel Lucent

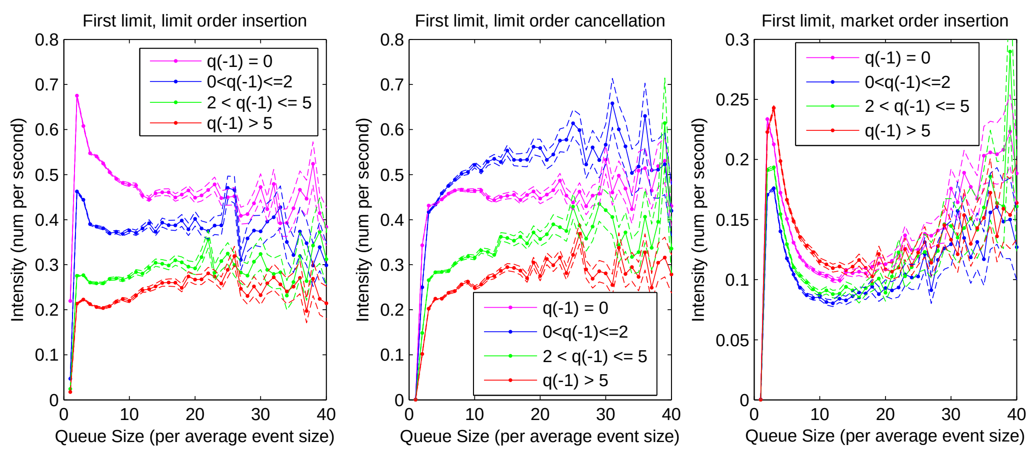

Figure I.15: Intensities at Q_1 as functions of $\mathscr{S}_{m,l}(q_{-1})$ and q_1, Alcatel Lucent

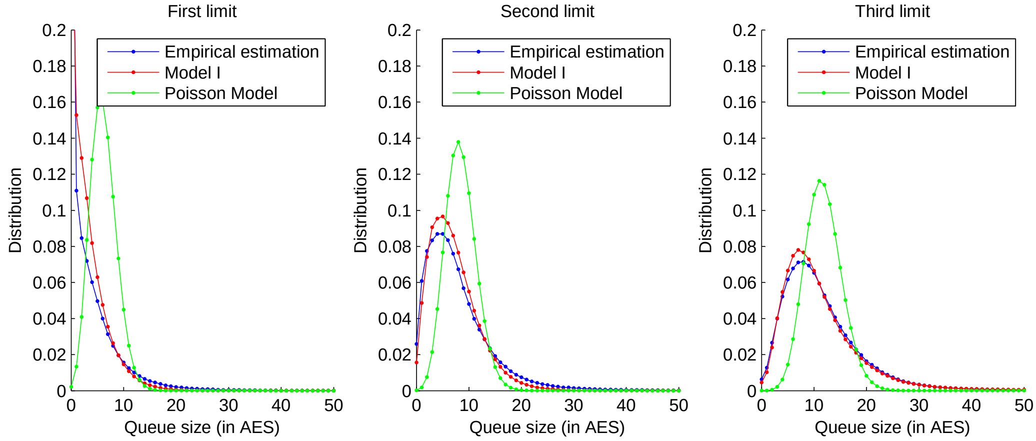

Figure I.16: Queue distribution, Alcatel Lucent

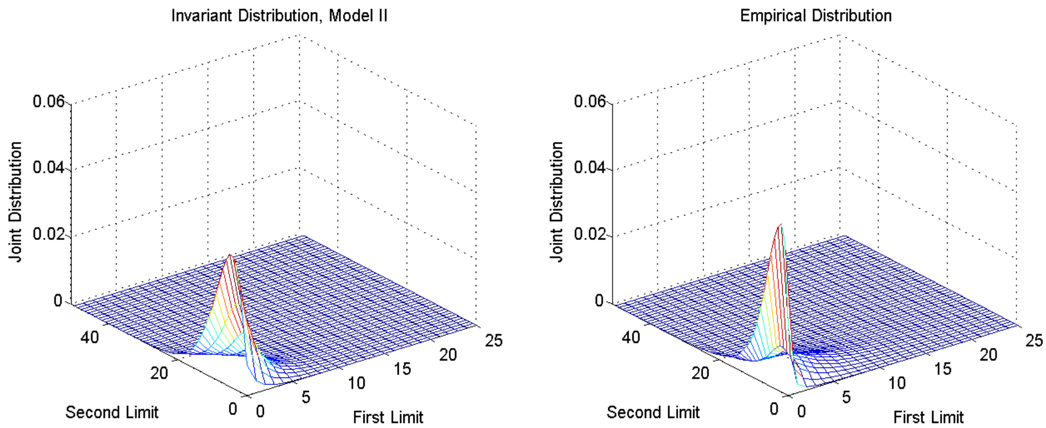

Figure I.17: Model II^a: joint distribution of q_1, q_2, Alcatel Lucent

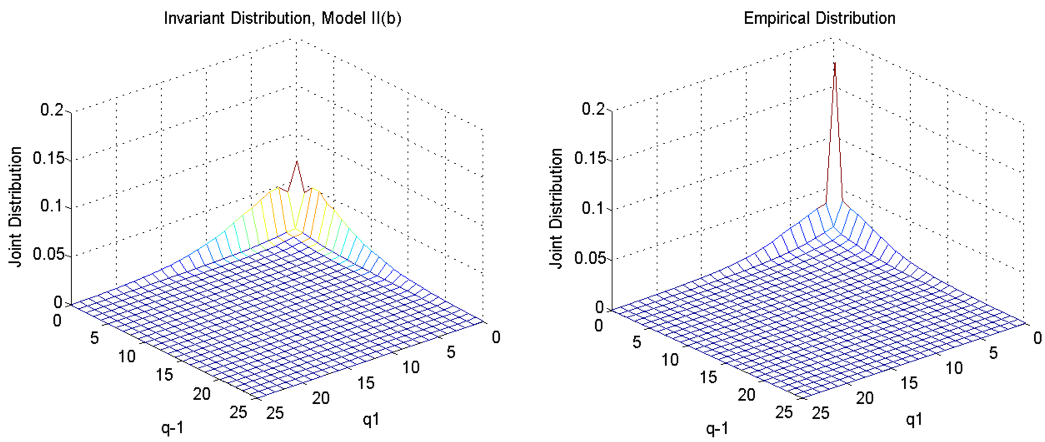

Figure I.18: Model II^b: joint distribution of q_{-1}, q_1, Alcatel Lucent the estimated values of AES at different distances to p_{ref} and the estimated value of ATS, for the stocks France Telecom and Alcatel-Lucent.

stock	ATS	AES_1	AES_2	AES_3
France Telecom	637	836	1068	1069
Alcatel Lucent	2340	3033	3451	3528

Table I.2: AES and ATS (in number of stocks)

CHAPTER II

A General Framework for Markovian Order Book Models

Abstract

We present a general Markovian framework for order book modeling. Through our approach, we aim at providing a tool enabling to get a better understanding of the price formation process and of the link between microscopic and macroscopic features of financial assets. To do so, we propose a new method of order book representation, and decompose the problem of order book modeling into two sub-problems: dynamics of a continuous-time double auction system with a fixed reference price; interactions between the double auction system and the reference price movements. State dependency is included in our framework by allowing the order flow intensities to depend on the order book state. Furthermore, contrary to most existing models, the impact of the order book updates on the reference price dynamics is not assumed to be instantaneous. We first prove that under general assumptions, our system is ergodic. Then we deduce the convergence towards a Brownian motion of the rescaled price process.

1 Introduction

Nowadays, most financial exchanges use a limit order book (LOB) mechanism. In these order-driven markets, market participants send their buy and sell orders to a continuous-time double auction system, where orders are matched according to their price and time priority. Understanding the LOB dynamics is one of the fundamental issues in the field of market microstructure and leads to many interesting applications in optimal execution, design of electronic trading algorithms, minimization of market impact costs, short-term predictions and regulation. In the recent years, many works have been devoted to the description of order book dynamics. Order book models can be essentially divided into two types: economic models, where one focuses on the behaviors of individual agents and their optimal decisions, see for example Parlour (1998), Foucault (1999) and Roşu (2009); statistical models, where the order flows are seen as random processes, see Smith, Farmer, Gillemot, and Krishnamurthy (2003), Cont, Stoikov, and Talreja (2010), Abergel and Jedidi (2011), Cont and De Larrard (2013), Lakner, Reed, and Stoikov (2013), Lachapelle, Lasry, Lehalle, and Lions (2013), Bayer, Horst, and Qiu (2014) and Abergel and Jedidi (2015). With the notable exception of Abergel and Jedidi (2015), where the authors consider the case of Hawkes-type dynamics, these models usually assume Poisson flows for the order arrival processes. Such assumption is mainly made for technical reasons, since it is well-known that it is not consistent with market data. In Huang, Lehalle, and Rosenbaum (2013), the authors propose to replace the Poisson assumption by a state-dependent approach where the intensities of the flows depend on the state of the LOB. This model, called "Queue-reactive" model, provides new insights for the order book dynamics, such as market participants behaviors conditional on different states of the order book, the LOB's asymptotic form and the bid-ask spread distribution. It is also a very relevant tool for practitioners in the perspective of transaction cost analysis of complex trading algorithms. In this paper, we aim at extending the Queue-reactive model to a more general framework, in which most of the existing statistical models can be included (up to minor modifications). Our goal is to give some theoretical results on the system's ergodicity as well as the asymptotic scaling limit of the price process.

In the LOB, price levels are discretized by a minimum price change unity called the tick value (denoted by α). Market participants can place their buy/sell orders at any level which is a multiple of the tick value and these orders will either stay in the LOB (a buy order with price lower than the current best ask price, or a sell order with price higher than the current best bid price, this type of orders being called "limit order"), or be matched with the existing orders in the LOB (this type of orders being called "market order"). The LOB, as its name suggests, is composed of all unmatched limit orders and can be seen as a (rough) approximation of the current microstructural supply and demand on the different price levels.

Current statistical models differ in their way of representing the LOB. In Cont, Stoikov, and Talreja (2010), the price grid is supposed to be finite $(n_{min}\alpha, \ldots, n_{max}\alpha)$, and the LOB is represented by a $n_{max} - n_{min} + 1$ dimensional vector that records the buying/selling quantities at each of these price levels. In such representation, the different limits are indexed by their absolute price level. In practice, to cover the intra-day price range $[p_{min}, p_{max}]$, the dimension of the state space have to be at least $\frac{p_{max}-p_{min}}{\alpha} + 1$[1], which is typically a very large number. Another way of representing the order book state is to use the relative indexing method. Following ideas from the Zero-intelligence model of Smith, Farmer, Gillemot, and Krishnamurthy (2003), Abergel and Jedidi (2011) propose to use the best bid and the best ask prices as two reference prices to index the limits. In that case, the LOB is made of the following elements: the two reference prices $p_{bestbid}$ and $p_{bestask}$, and the limits around them which are two K dimensional vectors a (for the ask side) and b (for the bid side). The vector $a = [a_1, \ldots, a_K]$ records the limit sizes at the price levels $[p_{bestbid} + \alpha, \ldots, p_{bestbid} + K\alpha]$, while the vector $b = [b_1, \ldots, b_K]$ records the limit sizes at the price levels $[p_{bestask} - \alpha, \ldots, p_{bestask} - K\alpha]$[2]. In practice, observing a market depth of five ticks is usually considered enough for most trading purposes. Consequently, for a typical stock with spread size of order five ticks, the value of K should be generally of order 10 so that essential information from the LOB is captured. Thus, the use of these two reference prices reduces significantly the dimensionality of the state space. Note that in this representation, the index of a limit at a given price level is no longer constant. Therefore, appropriate boundary conditions must be defined to deal with price changes.

In this paper, we propose an original representation of the order book, using only one reference price which is not necessarily directly observable from the order book state. We view this reference price as sort of market consensus about the underlying "efficient" price used by market participants when making their trading decisions. We keep K limits on each side of the reference price and the LOB is fully described by a $2K + 1$ dimensional vector, which is then modeled by a continuous-time Markov jump process. The use of this unique reference price gives us a lot of flexibility when modeling the order book. Since the reference price is no

[1]Of course if we choose $n_{min} = 1$ and $n_{max} = \infty$, we have a complete description of all the available buying/selling offers in the whole price grid $\alpha\mathbb{N}$.

[2]Note that in this representation, there can be overlaps between the vector a and b when the bid-ask spread is larger than one tick.

longer directly determined by the order book state, we can differentiate two types of jumps in the Markov process: pure order book state jumps (jumps for which the reference price stays constant) and common jumps (jumps for which a reference price change is involved). For large tick assets[3], such decomposition is proved to be very relevant when studying the conditional dependences between the dynamics of the LOB and its state, see Huang, Lehalle, and Rosenbaum (2013). Moreover, in this framework, we are able to easily incorporate exogenous price movements into the order book dynamics. This can be simply done adding a reference price jump component which is independent of the order book state.

At the high frequency scale, the LOB state is one of the two public information that are accessible to traders and their automates (the other being the history of the order flows). Thus it plays a very important role in their trading decisions. In our framework , the LOB is assumed to be a continuous-time Markov jump process, and the influence of the LOB state on the incoming flows is modeled through a state-dependent infinitesimal generator matrix for the jump process. Indeed, in practice, traders essentially rely on information deduced from the current LOB state when deciding to send an order at a specific price level. Various simplifying assumptions on the information set used by traders can be considered in our framework in order to facilitate the empirical studies. The index of a limit, for example, is probably one of the most important elements in their decision process, as it gives the distance between the target price and the reference price. Influence of other variables, such as the target limit's size, its relative distance to the current best offer queues and the size of its opposite queue is studied in Huang, Lehalle, and Rosenbaum (2013), and are shown to also have non-negligible effects on the dynamics of the order flows.

Under appropriate assumptions, the Queue-reactive model can be easily estimated using empirical data. It provides many interesting new insights on the origin of some micro-structural properties, such as the stylized empirical distribution of the LOB state. It has been shown in Bouchaud, Mézard, Potters, et al. (2002) that there exist some regularities in the order book's empirical form, that is the average value of the LOB state (a $2K$ dimensional vector in our model). From a theoretical point of view, these regularities are closely linked with the notion of ergodicity of the LOB system (the exact definition of ergodicity will be given in Section 3): Ergodicity ensures the convergence of the LOB state distribution towards an invariant probability measure. Thus the stylized form observed on market data might be explained by a law of large number type phenomenon for this invariant distribution. This hypothesis is supported by empirical studies in Huang, Lehalle, and Rosenbaum (2013), in which the authors compare the theoretical asymptotic distributions in our model with empirical estimations, and show that they are very close. In Huang, Lehalle, and Rosenbaum (2013), some assumptions are made to ensure the LOB system's ergodicity. In this paper, we want to generalize them in an extended framework where the volume of the orders is no longer constant and the influence of the order book state on the dynamics of the reference price may not be instantaneous.

Another important element in order book modeling is the asymptotic behavior of the price. Such analysis is very relevant as it provides useful insights on the price formation process, and links the dynamics at the microscopic level with macroscopic features of the asset, such as its volatility. We prove that in our framework, the rescaled reference price process converges to a Brownian motion. An expression for the macroscopic volatility in terms of the flow rates is derived using a functional central limit theorem together with the strong mixing property of the price increments, in the spirit of Abergel and Jedidi (2011).

[3] A large tick asset is defined as an asset whose bid-ask spread is almost always equal to one tick.

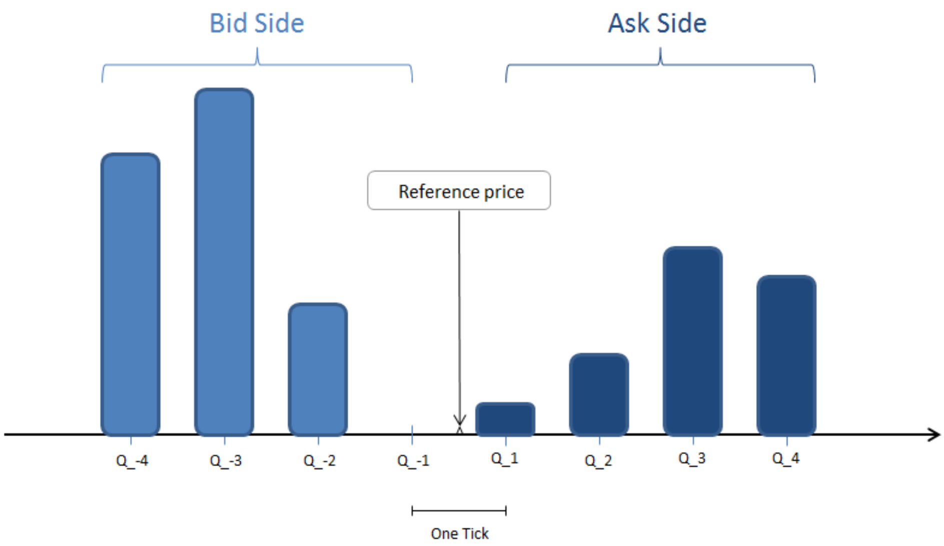

Figure II.1: Limit order book

The paper is organized as follows. In Section 2, we set up the general Markovian framework. Section 3 discusses the ergodicity properties of the model. The diffusive limit of the rescaled price process is stated in Section 4. Finally, in Section 5, we give some specific models which can be seen as particular cases of our framework. The technical proofs are gathered in an appendix.

2 A general Markovian framework

2.1 Representation of the order book

In our framework, the order book is made of two elements: the center position: a certain reference price (note that here the center position is not necessarily the mid price) and the shape of the book (the queue sizes around the reference price), see Figure II.1 for an example. The center position, denoted by $p_{ref} \in \{n\alpha + 0.5\alpha, n \in \mathbb{Z}\}$[4], can be seen as the current consensus price level and is used to index the limits. We write $Q_{\pm i}$ for the queue at the price level $p_{ref} \pm (i - 0.5)\alpha$ and denote its size by $q_{\pm i}$ and its price by p_i. Then the total order book's shape at time t is an infinite vector $q(t) = [..., q_{-k}(t), ...q_{-1}(t), q_1(t), ..., q_k(t), ...]$, with $q_i \in \mathbb{Z}$, and $|q_i|$ is the number of orders at the limit Q_i ($q_i < 0$ if these orders are bid orders and $q_i > 0$ if they are ask orders. Note that in such a representation, one can have $q_i \geq 0$ or $q_i < 0$ for all i). The LOB information at time t is therefore fully represented by $X(t) := (q(t), p_{ref}(t))$, $t \geq 0$. To restrict the value of $X(t)$, we consider only K limits on each side, and thus have now $q(t) = [q_{-K}(t), ...q_{-1}(t), q_1(t), ..., q_K(t)]$ and $X(t) \in \mathbb{Z}^{2K} \times \alpha(0.5 + \mathbb{Z})$.

The state space of $q(t)$ is actually smaller than $\mathbb{Z}^{2K}$. More specifically, let us define

$$\begin{aligned} i_{bestbid}(q) &= \max(-K-1, \sup\{i | q_i < 0\}) \\ i_{bestask}(q) &= \min(K+1, \inf\{i | q_i > 0\}). \end{aligned}$$

[4]For the generality of the framework, we allow for negative prices, however, in practice, the model should of course be used over a reasonably small time interval so that prices remain positive.

The state space Ω of $q(t)$ is defined as the set of all $q \in \mathbb{Z}^{2K}$, such that: for all $i \in \{-K, \ldots, K\}$, $q_i \leq 0$ if $i \leq i_{bestbid}(q)$ and $q_i \geq 0$ if $i \geq i_{bestask}(q)$.

2.2 Dynamics of the order book

In our general framework, we model the LOB vector $X(t)$ as a continuous-time Markov jump process, whose infinitesimal generator matrix $\mathcal{Q}$ will be given in Equation (II.4). We differentiate two types of jumps in the order book dynamics: pure order book state jumps (for which p_{ref} stays constant) and common jumps (for which a change in the value of p_{ref} is involved).

Pure order book state jump

There are three[5] types of orders that interact directly with the order book and trigger pure order book state jumps:

- Limit orders: insertion of a new order in the order book (a buy order at a lower price than the best ask price, or a sell order at a higher price than the best bid price).
- Cancellation orders: cancellation of an already existing order in the order book.
- Market orders: consumption of available liquidity (a buy or sell order at the best available price).

In the seminal work of Smith, Farmer, Gillemot, and Krishnamurthy (2003), the arrival times of the above three types of orders at different price levels are assumed to be mutually independent and exponentially distributed. Furthermore, each order has unit size. In our approach, the size of the jumps, which represents the amount of volume inserted to/removed from the LOB for a given event, is random. Moreover, the arrival rate of a given jump is assumed to be function of the index of the target price, the current LOB state vector $q(t)$, the direction of the jump and its size. That is, for any $q, q' \in \Omega$ $(q \neq q')$, $p \in \alpha(0.5 + \mathbb{Z})$, $n \in \mathbb{N}^+$ and any $e_i = (a_{-K}, \ldots, a_i, \ldots, a_K)$ ($a_j = 0$ for $j \neq i$ and $a_i = 1$), we have in cases where $q + ne_i \in \Omega$ and $q - ne_i \in \Omega$:

$$\begin{aligned}
\mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q,n) \\
\mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q,n) \\
\tilde{\mathcal{Q}}_{(q,p),(q',p)} &= 0, \text{ otherwise,}
\end{aligned} \tag{II.1}$$

where the f_i and g_i are $2K$ functions: $\Omega \times \mathbb{N}^+ \to \mathbb{R}^+$.
Note that in (II.1), $f_i(q,n)$ and $g_i(q,n)$ have different meanings for different i and q. For example, when $i \geq i_{bestask}$, $f_i(q,n)$ represents the arrival rate of sell limit orders of size n, and $g_i(q,n)$ the sum of the rate of cancellations of size n and the arrival rate of market buy orders of size n. When $i \leq i_{bestbid}$, the role of $f_i(q,n)$ and $g_i(q,n)$ are switched. Note also that $q \pm ne_i$ is not always in the state space Ω even when $q \in \Omega$. Thus some values of the functions f_i and g_i are not needed in Equation (II.1) and assumed to be equal to zero. Furthermore, so that there is no absorbing state, we assume

$$\sum_i \sum_n \big(f_i(q,n) + g_i(q,n)\big) > 0.$$

[5] Four if we also consider the modification orders. We view modification orders as a combination of a cancellation and an insertion order that arrive in a very short time interval.

Common jumps

The reference price p_{ref} can be viewed as a consensus value on the “efficient" price and takes discretized values in $\alpha(0.5+\mathbb{Z})$. In practice, this reference price is built based on two sets of information: the current state of the LOB and the historical order flows. We have in mind that p_{ref} moves in a Markovian manner, so its dynamics depends on the present information only, that is the current state of the LOB. In our framework, we restrict the price jump size to one tick at each time. We use two functions $u, d : \Omega \to \mathbb{R}^+$ to describe respectively the rate of positive and negative jumps:

$$\begin{aligned}
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p+\alpha)} &= u(q) \\
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p-\alpha)} &= d(q) \\
\sum_{q'\in\Omega} \mathcal{Q}_{(q,p),(q',p\pm n\alpha)} &= 0, \text{ for } n \geq 2.
\end{aligned} \tag{II.2}$$

To understand Equation (II.2), let us first consider the following simple example where the LOB state information is summarized by the difference between the current value of p_{ref} and the mid price p_{mid}:

Example 1.

$$\begin{aligned}
i_{mid} &= (i_{bestbid} + i_{bestask})/2, \\
u(q) &= \theta_0 + \theta_1 \max(0, \alpha(i_{mid} - 0.5)) \\
d(q) &= \theta_0 + \theta_1 \max(0, -\alpha(i_{mid} + 0.5)),
\end{aligned}$$

with $\theta_0 \geq 0$ and θ_1 a positive constant representing the intensity of the adjustment of p_{ref} towards p_{mid}.

In the above example, we assume that the reference price jump rate depends on the deviation of the current value of p_{ref} from p_{mid}. Indeed, p_{mid} is often considered an approximation of the LOB center implied by its current state. We may also include other LOB information such as the available quantities at $Q_{i_{bestbid}}$ and $Q_{i_{bestask}}$ when defining $u(q)$ and $d(q)$. Such additional variables increase the complexity of our model but make it more realistic.

Remark 1. *In the already mentioned Queue-reactive model, changes of p_{ref} are triggered by the order book events that modify p_{mid}, while in Equation* (II.2)*, they are driven by the order book state. Although the Queue-reactive model cannot exactly be seen as a particular case of the framework presented in this paper (see Section 5), most of the theoretical results shown in Sections 3 and 4 still hold (with some minor modifications in some assumptions). In particular, one can prove the diffusive limit of the reference price in the Queue-reactive model applying almost the same method as that used here.*

As soon as p_{ref} changes, the value of q_i switches immediately to the value of one of its neighbors (right if p_{ref} increases, left if it decreases). As we keep only K limits on each side, two boundary distributions π_{-K} and π_K are introduced for generating the new queue sizes at Q_{-K} (when p_{ref} decreases) and Q_K (when p_{ref} increases). To possibly incorporate external information, we moreover assume that with probability θ^{reinit}, the LOB state vector $q(t)$ is redrawn from some distribution (π^{inc} if p_{ref} increases, π^{dec} if p_{ref} decreases) when p_{ref} changes. As shown in Huang, Lehalle, and Rosenbaum (2013), models where price dynamics are purely endogenous, driven by order flows only, are usually not able to reproduce some of the important macroscopic features of prices, such as the volatility. Thus the parameter θ^{reinit} can be understood as the percentage of price changes due to exogenous information, in which case market participants readjust very quickly their order flows around the new reference price, as if a new state of the LOB was drawn from its (invariant) distribution (ergodicity conditions are discussed in the next section).

For $q \in \Omega$, we write $q^+ = [q_{-K+1}, ..., q_{-1}, q_1, ..., q_K]$, $q^- = [q_{-K}, ..., q_{-1}, q_1, ..., q_{K-1}]$, $[q^+, l] = [q_{-K+1}, ..., q_{-1}, q_1, ..., q_K, l]$ and $[l, q^-] = [l, q_{-K}, ..., q_{-1}, q_1, ..., q_{K-1}]$. Under the above assumptions, we have for $l \in \mathbb{Z}$ and $q, q', q'' \in \Omega$ such that $q'^+ \neq q^+$ and $q''^- \neq q^-$:

$$
\begin{aligned}
\mathcal{Q}_{(q,p),([q^+,l],p+\alpha)} &= (1-\theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+,l]) \\
\mathcal{Q}_{(q,p),(q',p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\
\mathcal{Q}_{(q,p),([l,q^-],p-\alpha)} &= (1-\theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l,q^-]) \\
\mathcal{Q}_{(q,p),(q'',p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q'').
\end{aligned} \tag{II.3}
$$

The infinitesimal generator matrix of $X(t)$

Equations (II.1), (II.2) and (II.3) give a complete description of the infinitesimal generator matrix $\mathcal{Q}$ of the process $X(t)$, which is summarized in the following assumption.

Assumption 21. *Let $q, q', q'', \tilde{q} \in \Omega$, $p, \tilde{p} \in \alpha(0.5+\mathbb{Z})$, $n \in \mathbb{N}^+$, $l \in \mathbb{Z}$ be such that $q'^+ \neq q^+$ and $q''^- \neq q^-$. The process $X(t)$ is an irreducible Markov jump process with aperiodic embedded chain whose infinitesimal generator matrix $\mathcal{Q}$ is of the following form (with $2K$ functions $f_i, g_i : \Omega \times \mathbb{N}^+ \to \mathbb{R}^+$ and two functions $u, d : \Omega \to \mathbb{R}^+$) :*

$$
\begin{aligned}
\mathcal{Q}_{(q,p),(q+ne_i,p)} &= f_i(q,n) \\
\mathcal{Q}_{(q,p),(q-ne_i,p)} &= g_i(q,n) \\
\mathcal{Q}_{(q,p),([q^+,l],p+\alpha)} &= (1-\theta^{reinit})u(q)\pi_K(l) + \theta^{reinit}u(q)\pi^{inc}([q^+,l]) \\
\mathcal{Q}_{(q,p),(q',p+\alpha)} &= \theta^{reinit}u(q)\pi^{inc}(q') \\
\mathcal{Q}_{(q,p),([l,q^-],p-\alpha)} &= (1-\theta^{reinit})d(q)\pi_{-K}(l) + \theta^{reinit}d(q)\pi^{dec}([l,q^-]) \\
\mathcal{Q}_{(q,p),(q'',p-\alpha)} &= \theta^{reinit}d(q)\pi^{dec}(q'') \\
\mathcal{Q}_{(q,p),(q,p)} &= -\sum_{(\tilde{q},\tilde{p})\in\Omega\times\alpha(0.5+\mathbb{Z}),(\tilde{q},\tilde{p})\neq(q,p)} \mathcal{Q}_{(q,p),(\tilde{q},\tilde{p})}, \\
\mathcal{Q}_{(q,p),(\tilde{q},\tilde{p})} &= 0, \text{ otherwise.}
\end{aligned} \tag{II.4}
$$

Note that under Assumption (21), the dynamics of the process $X(t)$ is invariant under translations of the LOB center position: its infinitesimal matrix generator $\mathcal{Q}$ satisfies:

$$
\mathcal{Q}_{(q^1,p^1),(q^2,p^1+\beta)} = \mathcal{Q}_{(q^1,p^2),(q^2,p^2+\beta)},
$$

for any $q^1, q^2 \in \Omega$, $p^1, p^2 \in \alpha(0.5+\mathbb{Z})$ and $\beta \in \alpha\mathbb{Z}$. One can also remark that in our framework, the order book state process $q(t)$ alone is also a continuous-time Markov jump process, whose ergodicity is discussed in the next section.

2.3 Comparison with existing models

The first major difference between our approach and the existing Markovian models in the literature is the introduction of state dependency in the order book dynamics. Most of the current order book models follow the "Zero-intelligence" framework, using Poisson flows for the processes of order arrivals. The Poisson assumption is clearly unrealistic, see for example the empirical results in Huang, Lehalle, and Rosenbaum (2013). In our framework, we propose to incorporate the strategic behaviors of market participants via a mean-field game approach, assuming their decisions depend on an underlying "efficient" price p_{ref} and on the LOB state vector $q(t)$. Note also that Equation (II.1) allows us to have jumps of random size in the order book's shape, while a constant jump size is often assumed in the other existing models.

We also introduce a new method of LOB representation using one unique reference price p_{ref}. Most models use the best bid and best ask prices as two reference prices for indexing the buy and sell limits. In such models, these reference prices are directly determined by the order book state. In particular, changes in the order book state are immediately carried on the values of these prices. In our framework, p_{ref} is not necessarily deduced from the order book state. Therefore, we can assume that changes in the order book state affect the value of p_{ref} with some delay rate (the functions u and d introduced in Equation (II.2)). Thanks to this original representation, we can naturally decompose the order book dynamics into two parts: a continuous-time multidimensional queuing system (Equation (I.1)) and the dynamics of its center, that is the reference price (Equations (II.2) and (II.3)). Compared with the Queue-reactive model introduced in Huang, Lehalle, and Rosenbaum (2013), p_{ref} is no longer constrained within the bid-ask spread. This desirable feature gives us the possibility of separating the exogenous and endogenous parts in the price dynamics by choosing appropriate price jump rate functions u and d. For example, with the functions u and d of Example 1, one can interpret θ_0 as the exogenous part in the dynamics of p_{ref} and θ_1 as the intensity of the endogenous effects driving p_{ref} towards the current mid price level.

3 Ergodicity

In this section, we discuss ergodicity properties in our framework. To do so, we make some additional assumptions on the functions f_i, g_i, u and d.

Let
$$P_t(x,A) := \mathbb{P}[Y(t) \in A | Y(0) = x]$$
be the transition probability at time t of a continuous or discrete-time Markov process Y with state space $\mathcal{Y}$. In this work, we say that the process Y is *V-uniformly ergodic* if there exists a coercive[6] function $V > 1$, an invariant distribution π, $r \in (0,1)$ and $R > \infty$ such that for any $x \in \mathcal{Y}$ and $t \in \mathbb{R}^+$ (or $\mathbb{N}^+$ in discrete-time),

$$||P_t(x,.) - \pi(.)||_V \leq R r^t V(x), \tag{II.5}$$

where we write $||.||_V$ for the V-norm of a signed measure, see Meyn and Tweedie (1993, 2009). In continuous time, the main idea to prove such property is to design an appropriate Lyapunov function V: $\mathcal{Y} \to (1,\infty)$, on which the following negative drift condition is satisfied for some

[6] $|V(x)| \to +\infty$ as $||x|| \to \infty$.

$\gamma > 0$ and $B > 0$:

$$\begin{aligned}\mathcal{Q}V(y) &:= \sum_{y' \neq y} \mathcal{Q}_{yy'}[V(y') - V(y)] \\ &\leq -\gamma V(y) + B.\end{aligned}$$

Then by Theorem 6.1 in Meyn and Tweedie (1993), the Markov process Y is non-explosive and V-uniformly ergodic. Furthermore, by Theorem 4.2 in Meyn and Tweedie (1993) it is positive Harris recurrent. Note that the same kind of method is used in Abergel and Jedidi (2011) in order to show ergodicity properties of Zero-intelligence models.

As mentioned in the introduction, the LOB's ergodicity implies here the existence of a unique invariant distribution for the state vector q. This is relevant for explaining the stylized empirical distribution of the LOB state. Mostly, as we will see in Section 4, the ergodicity analysis is the basis for proving the diffusive limit of the reference price process.

3.1 When p_{ref} stays constant

We first discuss the V-uniform ergodicity of the process $q(t)$ when assuming $u(q) = d(q) = 0$ in Equation (II.4). Recall that the unused values of $f_i(q,n)$ and $g_i(q,n)$ in the definition of the queue dynamics, that is when $q \pm ne_i \notin \Omega$, are set to zero. With the convention $0/0 = 0$, we define

$$\begin{aligned}f_i^*(q) &:= \sum_n f_i(q,n) \\ g_i^*(q) &:= \sum_n g_i(q,n) \\ l_i(q,n) &:= \frac{f_i(q,n)}{f_i^*(q)} \\ k_i(q,n) &:= \frac{g_i(q,n)}{g_i^*(q)} \\ G^{f,i,q}(z) &:= \sum_{n=1}^{\infty} z^n l_i(q,n) \\ G^{g,i,q}(z) &:= \sum_{n=1}^{\infty} z^n k_i(q,n).\end{aligned}$$

Thus, when $f_i^*(q) > 0$ (resp. $g_i^*(q) > 0$), $l_i(q,.)$ (resp. $k_i(q,.)$) is a probability measure on $\mathbb{N}^+$ with moment-generating function $G^{f,i,q}(z)$ (resp. $G^{g,i,q}(z)$). We make the four following assumptions.

Assumption 22. *For any order book state q and any $i \geq i_{bestask}$, $g_i(q,n) = 0$ for any $n > q_i$ and for any order book state q and any $i \leq i_{bestbid}$, $f_i(q,n) = 0$ for any $n > -q_i$.*

Assumption 23. *There exists $z^* > 1$ such that for any q and i, $G^{f,i,q}(z^*) < \infty$ and $G^{g,i,q}(z^*) < \infty$. Furthermore, there exists $L > 0$ such that for any i,*

$$\overline{\lim_{z \to 1^+}} \sup_q [f_i^*(q) G^{f,i,q}(z) \mathbf{1}_{i > i_{bestbid}} + g_i^*(q) G^{g,i,q}(z) \mathbf{1}_{i < i_{bestask}}] < L.$$

Assumption 24. *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned}\overline{\lim_{z \to 1^+}} \sup_{(q,i): q_i > U, i \geq i_{bestask}} [f_i^*(q) - g_i^*(q) \frac{1 - G^{g,i,q}(z^{-1})}{G^{f,i,q}(z) - 1}] &< -r \\ \overline{\lim_{z \to 1^+}} \sup_{(q,i): q_i < -U, i \leq i_{bestbid}} [g_i^*(q) - f_i^*(q) \frac{1 - G^{f,i,q}(z^{-1})}{G^{g,i,q}(z) - 1}] &< -r. \qquad \text{(II.6)}\end{aligned}$$

Assumption 25. *For any $z > 1$,*

$$\begin{aligned} B_f(z) &:= \inf_{(q,i):q_i>U, i\geq i_{bestask}} (G^{f,i,q}(z)-1) > 0 \\ B_g(z) &:= \inf_{(q,i):q_i<-U, i\leq i_{bestbid}} (G^{g,i,q}(z)-1) > 0. \end{aligned}$$

To understand the practical meaning of these assumptions, let us consider the following example where the pure order book state jumps are assumed to have constant size equal to one. In such situation, the four assumptions above can be rewritten as follows and are much easier to interpret.

Example 2. *LOB model with constant order size.*

- *For $n \geq 2$, $f_i(q,n) = g_i(q,n) = 0$ for any $q \in \Omega$.*
- *There exists $L > 0$ such that for any $i \in \{-K,...,K\}$ and $q \in \Omega$,*

$$f_i(q,1)\mathbf{1}_{i>i_{bestbid}} + g_i(q,1)\mathbf{1}_{i<i_{bestask}} < L.$$

- *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned} \sup_{(q,i):q_i>U, i\geq i_{bestask}} [f_i(q,1) - g_i(q,1)] &< -r \\ \sup_{(q,i):q_i<-U, i\leq i_{bestbid}} [g_i(q,1) - f_i(q,1)] &< -r. \end{aligned}$$

Basically, Assumption 22 says that a bid/ask limit cannot become an ask/bid limit in a single queue update event, that is the queue size cannot revert its sign in a single jump[7]. From Example 2, we see that Assumption 23 essentially states that the total intensity of the order insertion processes in the bid and ask side remains uniformly bounded with respect to the state of the LOB. Assumption 24, which is the most important for the system's ergodicity, forces the individual queue sizes $|q_i|$ to decrease when they become larger than a certain threshold. From these assumptions, we obtain the following theorem proved in appendix for the Markov process $q(t)$.

Theorem 1. *When $u = d = 0$, under Assumptions 21, 22, 23, 24 and 25, the continuous-time Markov jump process $q(t)$ is non-explosive, V-uniformly ergodic and positive Harris recurrent.*

Consider now the embedded Markov chain $q(n)$ defined by $q(n) = q(J_n)$, where J_n is the time of the n-th jump, and $q(J_n)$ the state of the LOB after this event. The study of the embedded Markov chain is an important step in order to obtain the diffusivity of the price process in our setting. We write

$$a_i^*(q) = \frac{f_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]}, \; b_i^*(q) = \frac{g_i^*(q)}{\sum_i [f_i^*(q) + g_i^*(q)]},$$

for the proportions of queue size increases and decreases, and replace Assumption 24 by the following one.

[7] This assumption is not really mandatory but is realistic and technically quite convenient.

Assumption 26. *There exist $r > 0$ and $U > 1$ such that*

$$\begin{aligned}
\overline{\lim_{z\to 1^+}} \sup_{(q,i):q_i>U, i\geq i_{bestask}} [a_i^*(q) - b_i^*(q)\frac{1-G^{g,i,q}(z^{-1})}{G^{f,i,q}(z)-1}] &< -r \\
\overline{\lim_{z\to 1^+}} \sup_{(q,i):q_i<-U, i\leq i_{bestbid}} [b_i^*(q) - a_i^*(q)\frac{1-G^{f,i,q}(z^{-1})}{G^{g,i,q}(z)-1}] &< -r.
\end{aligned}$$

The following theorem is proved in appendix.

Theorem 2. *When $u = d = 0$, under Assumptions 21, 22, 23, 25 and 26, the embedded discrete-time Markov chain $q(n)$ is V-uniformly ergodic and positive Harris recurrent.*

3.2 General case

We are now interested in the case where u and d are no longer fixed to 0. Recall that $q(n)$ represents the state of the LOB after the n-th event and $p_{ref}(n)$ is the reference price (the center of the LOB) after the n-th event. We thus consider here the process of reference price increments $c(n)$ (since the reference price itself is of course not ergodic), defined as the reference price change at the n-th event:

$$c(n) = p_{ref}(n) - p_{ref}(n-1),$$

and the embedded chain $Y(n) = (q(n), c(n))$, $n \in \mathbb{N}$, with $c(0)$ an artificial starting value. The process $Y(n)$ remains obviously Markovian. For some $z > 1$, let

$$V^z([q,c]) = \sum_{i=-K, i\neq 0}^{K} z^{|q_i|-U}.$$

We make two additional assumptions for the general case.

Assumption 27. *There exist $z > 1$ and $L^\pi > 0$ such that for Q^{inc}, Q^{dec}, Q_K, Q_{-K} four random variables such that $Q^{inc} \sim \pi^{inc}$, $Q^{dec} \sim \pi^{dec}$, $Q_K \sim \pi_K$ and $Q_{-K} \sim \pi_{-K}$:*

$$\mathbb{E}[V^z([Q^{inc}, c])] + \mathbb{E}[V^z([Q^{dec}, c])] + \mathbb{E}[z^{|Q_K|-U}] + \mathbb{E}[z^{|Q_{-K}|-U}] \leq L^\pi.$$

Assumption 28. *There exists a finite set $W \subset \Omega$ such that the upper bound of the proportion of reference price jumps in any order book state q is smaller than one on Ω / W:*

$$\sup_{q\in\Omega/W} \frac{u(q)+d(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} < 1.$$

Assumption 27 is technical and imposes some regularities on the four distributions π^{inc}, π^{dec}, π_K and π_{-K}. Assumption 28 ensures that a reference price change is not the only possible event except for a finite number of LOB states. Under these assumptions, we have the following theorem proved in appendix on the ergodicity of the embedded chain $Y(n)$.

Theorem 3. *Under Assumptions 21, 22, 23, 25, 26, 27 and 28, the embedded discrete-time Markov chain $Y(n) = (q(n), c(n))$ is V-uniformly ergodic and positive Harris recurrent.*

4 Scaling limits

We are now interested in the scaling limit of the reference price process. Let J_i be the time of the i-th jump of the process. Let

$$N(t) = \inf\{n,\ J_n \leq t\}$$

be the number of events until time t, with the convention $\inf\{\emptyset\} = 0$. Let $Z(n)$ be the cumulative price change until the n-th event, that is $Z(0) = 0$ and for $n \geq 1$:

$$Z(n) = \sum_{i=1}^{n} c(i).$$

We have

$$Z(N(t)) = p_{ref}(t) - p_{ref}(0).$$

Thus it represents the reference price at time t recentered its starting value. We show in this section the diffusive behavior of $Z(N(t))$ as n tends to infinity.

Consider again the embedded chain $Y(n) = (q(n), c(n))$. From Theorem 3, $Y(n)$ is V-uniformly ergodic towards an invariant distribution π^*. We have the following theorem for the rescaled price process in event time $\hat{S}^{(n)}(t) := \frac{Z(\lfloor nt \rfloor)}{\sqrt{n}}$ (π^* is defined as the invariant distribution of the Markovian chain $Y(n)$).

Theorem 4. *Under Assumptions 21, 22, 23, 25, 26, 27 and 28, if* $\mathbb{E}_{\pi^*}[c(0)] = 0$, *then the series*

$$\sigma^2 = \mathbb{E}_{\pi^*}[c_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}_{\pi^*}[c_0 c_n], \tag{II.7}$$

converges absolutely, with π^* *the invariant distribution of* $(q(n), c(n))$. *Furthermore, if* $Y(0) \sim \pi^*$, *we have the following convergence in law in* $D[0,\infty)$*:*

$$\hat{S}^{(n)}(t) \overset{n\to\infty}{\to} \sigma B(t),$$

where $B(t)$ *is a standard Brownian motion.*

Proof. This theorem is a direct application of Theorem 19.1 in Billingsley (2009). Indeed, the sequence c_n is clearly stationary and ergodic in the sense of Billingsley (2009) (for example since it is stationary and mixing). Moreover, it has a finite second order moment and for all n, $\mathbb{E}_{\pi^*}[c_n] = \mathbb{E}_{\pi^*}[c_0] = 0$. □

Theorem 4 shows that in event time, the large scale limit of the reference price process is a Brownian motion. However, the most relevant question is that of the large scale limit of the reference price in calendar time. Thus we now consider the process

$$\tilde{S}^{(n)}(t) = \frac{Z\big(N(nt)\big)}{\sqrt{n}}.$$

To prove the diffusivity of $\tilde{S}^{(n)}(t)$, we need a last assumption which is a bound of the expected value on the waiting time between two events.

Assumption 29. *There exists some* $m > 0$, *such that*

$$\inf_{q\in\Omega} \big\{\sum_i (f_i^*(q) + g_i^*(q)) + u(q) + d(q)\big\} > m.$$

Let τ_n be the inter-arrival time between the n-th and the $(n-1)$-th jumps of the Markov process X. We then have the following theorem proved in appendix for the long term behavior of the reference price in calendar time.

Theorem 5. *Under Assumptions 21, 22, 23, 25, 26, 27, 28 and 29, the process* $(q(n), c(n), \tau(n))$ *is positive Harris recurrent. Furthermore, if* $\mathbb{E}_{\pi^*}[c(0)] = 0$ *and* $Y(0) \sim \pi^*$, *then*

$$\tilde{S}^{(n)}(t) \overset{n\to\infty}{\to} \frac{\sigma}{\sqrt{\mathbb{E}_{\pi^{**}}[\tau(1)]}} B(t),$$

with π^{**} *is the invariant distribution of* $(q(n), c(n), \tau(n))$ *and* σ *defined in* (II.7).

Theorem 5 discusses the scaling limit of the underlying reference price. However, the difference between this price and the more usual $p_{bestbid}(t)$, $p_{bestask}(t)$ or $p_{mid}(t)$ being bounded by $2K$, the same result applies replacing the reference price by any of those prices.

5 Some specific models

The Markovian setting proposed in this work allows us for a wide range of possibilities for modeling order book dynamics. The goal of this section is to give some natural and tractable examples of models, essentially already introduced in the literature, which can be seen as particular cases of our general framework. Together with the dynamics of the models, we provide sufficient conditions so that the assumptions made in the previous sections are satisfied in these specific models. Thus the ergodicity and diffusive scaling properties apply in all these models.

5.1 Best bid/best ask Poisson model (Cont and De Larrard (2013))

The basic idea of this first model, inspired by that introduced in Cont and De Larrard (2013), is to use a constant spread size (fixed to 1 tick) and to consider only two limits in the order book.

Example 3. *Poisson model with* $K = 1$.

- *We take* $K = 1$, $\theta^{reinit} = 1$ *and assume that the functions* f_i, g_i, u *and* d *have the following forms, with* $0 < \lambda < \mu < \infty$*:*

$$\begin{aligned} f_1(q,n) &= \lambda \mathbf{1}_{n=1} \\ g_1(q,n) &= \mu \mathbf{1}_{q_1>0} \mathbf{1}_{n=1} \\ f_{-1}(q,n) &= \mu \mathbf{1}_{q_{-1}<0} \mathbf{1}_{n=1} \\ g_{-1}(q,n) &= \lambda \mathbf{1}_{n=1} \\ u(q) &= \theta \mathbf{1}_{q_1=0} \\ d(q) &= \theta \mathbf{1}_{q_{-1}=0}. \end{aligned}$$

- π^{inc} *and* π^{dec} *satisfy Assumption 27 and*

for any $q_{-1} > 0$, $q_1 \in \mathbb{Z}$,
$$\pi^{inc}(q_{-1}, q_1) = \pi^{dec}(q_{-1}, q_1) = 0,$$
for any $q_1 < 0$, $q_{-1} \in \mathbb{Z}$,
$$\pi^{inc}(q_{-1}, q_1) = \pi^{dec}(q_{-1}, q_1) = 0.$$

Note that here, the boundary distributions π_K and π_{-K} are no longer needed, since the order book reinitialization probability θ^{reinit} is set to one.

In this model, the role of p_{ref} is very close to that of p_{mid}, which splits the order book into two parts: the bid side (Q_{-1}) and the ask side (Q_1). The limit order insertion, cancellation and market order insertion processes are assumed to be independent Poisson processes. The size of these orders is assumed to be constant and p_{ref} jumps with rate 0 when none of the queues $Q_{\pm 1}$ is empty, with rate θ to the right side when Q_1 is empty, with rate θ to the left side when Q_{-1} is empty. When the value of θ is very large, the price jump is almost instantaneous as soon as one of the two queues becomes empty. In that case, this model becomes very close to that proposed in Cont and De Larrard (2013), where an infinite rate is used (note that the convergence of the rescaled price process can still be proved with some minor modifications in such case of infinite jump rate).

5.2 Poisson model with $K > 1$

It is natural to try to extend the previous Poisson model in order to include more queues in the order book and to allow for a spread size different from one tick. In such model, the role of p_{ref} is slightly different since it is not necessarily the mid price. Again, p_{ref} can be understood here as the underlying efficient price that determines the order arrival intensities at different price levels. Now buy/sell limit orders can be inserted both on the right side and on the left side of p_{ref}.

Example 4. *Poisson model with $K > 1$.*

- *The functions f_i, g_i, u and d have the following forms, for $i = -K, \ldots, K$:*

$$\begin{aligned}
f_i(q,1) &= \lambda_i \mathbf{1}_{i > i_{bestbid}(q)} + \gamma_i \mathbf{1}_{i = i_{bestbid}(q)} \\
&\quad + \mu_i \mathbf{1}_{i \le i_{bestbid}(q)} \mathbf{1}_{q_i < 0} \\
g_i(q,1) &= \lambda_{-i} \mathbf{1}_{i < i_{bestask}(q)} + \gamma_{-i} \mathbf{1}_{i = i_{bestask}(q)} \\
&\quad + \mu_{-i} \mathbf{1}_{i \ge i_{bestask}(q)} \mathbf{1}_{q_i > 0} \\
f_i(q,n) &= 0, \textit{for } n > 1 \\
g_i(q,n) &= 0, \textit{for } n > 1 \\
u(q) &= \theta_{i_{bestask}(q)} \\
d(q) &= \theta_{-i_{bestbid}(q)}.
\end{aligned}$$

- *π^K, π^{-K}, π^{inc} and π^{dec} satisfy Assumption 27.*

- *For any $i, j \in \{-K, \ldots, K\}$, $i < j$, we have*

$$\begin{gathered}
\mu_{-i} > \lambda_i > 0 \\
0 \le \theta_i \le \theta_j.
\end{gathered}$$

Limit orders, cancellations and market orders (which consume the quantities at the best offer limits) are modeled by independent Poisson processes with different intensities $\lambda_i^{buy/sell}$, $\mu_i^{buy/sell}$ and $\gamma_i^{buy/sell}$, depending on the distance from their target price to p_{ref}. We assume bid-ask symmetry in this model, that is $\lambda_i^{buy} = \lambda_{-i}^{sell}$, $\mu_i^{buy} = \mu_{-i}^{sell}$ and $\gamma_i^{buy} = \gamma_i^{sell}$, thus we omit the index buy/sell in the above equations. Remark that the intensity of the buy/sell market order flow at the best limit γ_i is a function of i, that is the position of the best limit with respect to the reference price. This allows us to model the fact that market participants have different behaviors towards the best limit, depending on their evaluation of the reference price.

The reference price jump dynamics is modeled by a sequence of increasing rates $\theta_i, i \in \{-K, ..., K\}$. This means that the larger the index of the best ask queue, the larger the probability of p_{ref} to increase and the smaller the index of the best bid queue, the larger the probability of p_{ref} to decrease. Note that in this model, we no longer assume any specific value for the reinitialization probability θ^{reinit} and use Assumption 27 to impose some properties on the boundary distributions π^K and π^{-K} and the initialization distributions π^{inc} and π^{dec}.

5.3 Zero-intelligence model

We now present a different way of extending the Poisson model with $K = 1$ in order to include more queues in the order book. This modeling approach where two reference prices are used is called Zero-intelligence model and is introduced in Smith, Farmer, Gillemot, and Krishnamurthy (2003). It is also considered in Abergel and Jedidi (2011) and is the basis of Cont, Stoikov, and Talreja (2010). We define $\phi(i, j)$ as the absolute distance (in number of ticks) between the queue Q_i and Q_j and make the three following assumptions.

Example 5. *Zero-intelligence model.*

- *The functions f_i, g_i, u and d have the following forms, for $i = -K, ..., K$:*

$$\begin{aligned}
f_i(q,1) &= \lambda_{\phi(i,i_{bestbid}(q))}\mathbf{1}_{i>i_{bestbid}(q)} + \gamma\mathbf{1}_{i=i_{bestbid}(q)} + |q_i|\mu_{\phi(i,i_{bestask}(q))}\mathbf{1}_{i\le i_{bestbid}(q)} \\
g_i(q,1) &= \lambda_{\phi(i,i_{bestask}(q))}\mathbf{1}_{i<i_{bestask}(q)} + \gamma\mathbf{1}_{i=i_{bestask}(q)} + |q_i|\mu_{\phi(i,i_{bestbid}(q))}\mathbf{1}_{i\ge i_{bestask}(q)} \\
f_i(q,n) &= 0, \textit{for } n > 1 \\
g_i(q,n) &= 0, \textit{for } n > 1 \\
u(q) &= \theta_{i_{bestask}(q)} \\
d(q) &= \theta_{-i_{bestbid}(q)}.
\end{aligned}$$

- *π^K, π^{-K}, π^{inc} and π^{dec} satisfy Assumption 27.*

- *For any $i, j \in \{-K, ..., K\}$, $\lambda_{\phi(i,j)} > 0$ and $\mu_{\phi(i,j)} > 0$.*

- *For any $i, j \in \{-K, ..., K\}$, $i < j$, we have*

$$0 \le \theta_i \le \theta_j.$$

In this model, p_{ref} is no longer an underlying efficient price determining the order arrival intensities. These intensities now depend on the positions of two different prices: $p_{bestbid}$ and $p_{bestask}$. Limit orders, cancellations and market orders are still described by independent Poisson processes. Buy/sell limit orders are inserted in the queues to the left/right side of the best ask/best bid price, with intensities depending on the distance between their price level and the best ask/best bid price ($\lambda_{\phi(i,i_{bestask/bestbid})}$); cancellations of buy/sell orders are sent to the queues on the left/right side of the best ask/best bid price, with intensities being linear functions of the queue sizes ($|q_i|\mu_{\phi(i,i_{bestask/bestbid})}$); market buy/sell orders are sent to the best ask/best bid queue, with intensity γ. The reference price p_{ref} now provides the center of the $2K$ dimensional moving frame representing the LOB's state and the same modeling approach as in Section 5.2 is used for its dynamics.

5.4 Queue-reactive model (Huang, Lehalle, and Rosenbaum (2013))

In Huang, Lehalle, and Rosenbaum (2013), the Queue-reactive model for order books is introduced. This model takes into account the influence of the order book's state in determining the order arrival intensities (in a much more general way than considering only the position of the best bid and best ask queues). The Queue-reactive model assumes that no buy/sell limit order can be inserted on the right/left side of p_{ref} and uses the following assumption instead of Equation (II.2) for the dynamics of the jumps of p_{ref}.

Assumption 30. *Whenever p_{mid} increases (resp. decreases), p_{ref} increases (resp. decreases) by α with probability θ, provided that $q_1 = 0$ (resp. $q_{-1} = 0$) at that moment. Therefore, changes in the value of p_{ref} are possibly triggered by one of the three following events:*

- *The insertion of a buy (resp. sell) limit order within the bid-ask spread while Q_1 (resp. Q_{-1} is empty).*
- *The cancellation of the last limit order at one of the best offer queues.*
- *A market order which consumes the last limit order at one of the best offer queues.*

With some minor modifications in the proof of Theorem 5, one can prove that the scaling limit of p_{ref} in the Queue-reactive model is a Brownian motion. As explained above, in this model, changes of p_{ref} are triggered by events that modify the mid price. Here we propose a slightly modified version of the Queue-reactive model[8] by considering the following four assumptions (note that the state space Ω is reduced in that case to $\Omega^* := \{q \in \Omega, q_i \mathbf{1}_{i<0} \le 0, q_i \mathbf{1}_{i>0} \ge 0\}$).

Example 6. *Queue-reactive type model.*

- *The functions f_i, g_i, u and d have the following forms, for $i = -K, ..., K$:*

$$\begin{aligned}
f_i(q,1) &= \lambda_{|i|}(q_i)\mathbf{1}_{i>0} + \mu_{|i|}(-q_i)\mathbf{1}_{i<0}\\
g_i(q,1) &= \lambda_{|i|}(-q_i)\mathbf{1}_{i<0} + \mu_{|i|}(q_i)\mathbf{1}_{i>0}\\
f_i(q,n) &= 0, \textit{for } n > 1\\
g_i(q,n) &= 0, \textit{for } n > 1\\
u(q) &= \theta \mathbf{1}_{q_1=0}\\
d(q) &= \theta \mathbf{1}_{q_{-1}=0},
\end{aligned}$$

  with $\lambda_{|i|}$ and $\mu_{|i|}$ non-negative functions defined on $\mathbb{N}$, with $\mu_{|i|}(0) = 0$.

- *We have*

$$\sup_{i \in \{1,...,K\}, q_i \in \mathbb{N}} (\lambda_i(q_i)) < L < \infty.$$

- *There exist $r > 0$ and $U > 1$ such that for any $q_i > U$ and any $i \in \{1, ..., K\}$:*

$$\lambda_i(q_i) - \mu_i(q_i) \quad < \quad -r.$$

- *There exists $m' > 0$ such that for any $i \in \{1, ..., N\}$:*

$$\inf_{q_i \in \mathbb{N}} [\lambda_i(q_i) + \mu_i(q_i)] \quad > \quad m'.$$

[8] Model I in Huang, Lehalle, and Rosenbaum (2013) is used to describe the queue dynamics during constant reference price periods.

Compared with Assumption 30, changes in p_{ref} are now driven by the relative position of the mid price in the current order book state. Nevertheless, we can see that the two approaches are quite similar. In this model, p_{ref} always stays between $p_{bestask}$ and $p_{bestbid}$ (since $\mu_{|i|}(0) = 0$ implies that the queue sizes on the left/right side of p_{ref} never become positive/negative). Such model gives us a much larger choice on the intensity functions $\lambda_{|i|}$ and $\mu_{|i|}$ than when assuming Poisson flows. Furthermore, with enough data points, these functions can be estimated in a non-parametric way, as done in Huang, Lehalle, and Rosenbaum (2013). Finally the state-dependent approach provides us very interesting insights about the way the order book state influences market participants decisions and the mechanism making the empirical distribution of the order book arise from these decisions, see Huang, Lehalle, and Rosenbaum (2013).

6 Conclusion

In this work, we extend the order book modeling approach proposed in Huang, Lehalle, and Rosenbaum (2013) to a more general Markovian framework, allowing to take into account most relevant features of LOB dynamics such as:

- Dependencies between the order arrival processes and the LOB state.
- Endogenous movements of the underlying efficient price and influence of the LOB state on its dynamics.
- Exogenous movements of the underlying efficient price.
- Randomness in the size of the orders.

The ergodicity of the LOB system and the diffusive limit of the rescaled price process are established under general assumptions. Finally, to illustrate the usefulness and the relevance of our approach, several examples of classical models which can be seen as particular cases of our general framework are presented.

To get a fully satisfying model, a last step would probably be to allow for a non-Markovian component in the market order flow (since the Markov assumption is probably quite reasonable for the limit order and cancellation flows). This can for example be done using self-exciting processes, as in Abergel and Jedidi (2015). However, except for very specific cases (exponential Hawkes processes for example), adding such non-Markovian component would certainly require revising completely the mathematical approach to the model.

Ackowledgements

We thank Frédéric Abergel for helpful discussions.

7 Appendix

7.1 Proof of Theorem 1

Let us denote by $\tilde{\mathscr{Q}}$ the infinitesimal generator matrix of $q(t)$ when $u = d = 0$. The infinite matrix $\tilde{\mathscr{Q}}$ has the following form:

$$\begin{aligned}
\tilde{\mathscr{Q}}_{q,q+ne_i} &= f_i(q,n)\\
\tilde{\mathscr{Q}}_{q,q-ne_i} &= g_i(q,n)\\
\tilde{\mathscr{Q}}_{q,q} &= -\sum_{q\in\Omega,q'\neq q}\tilde{\mathscr{Q}}_{q,q'}\\
\tilde{\mathscr{Q}}_{q,q'} &= 0, \text{otherwise}.
\end{aligned}$$

For some $1 < z \leq z^*$ (recall that z^* is defined in Assumption 23), let us consider the function

$$V(q) = \sum_{i=-K,i\neq 0}^{K} z^{|q_i|-U}.$$

Since $q_i \geq 0$ for $i \geq i_{bestask}$, $q_i \leq 0$ for $i \leq i_{bestbid}$ and $q_i = 0$ for $i \in (i_{bestbid}, i_{bestask})$, we have

$$\begin{aligned}
\tilde{\mathscr{Q}}V(q) &:= \sum_{q'\neq q}\tilde{\mathscr{Q}}_{qq'}[V(q') - V(q)]\\
&= \sum_{i\leq i_{bestbid}}\sum_{n=1}^{\infty}[(z^{-q_i+n-U} - z^{-q_i-U})g_i(q,n) + (z^{|q_i+n|-U} - z^{-q_i-U})f_i(q,n)]\\
&+ \sum_{i\geq i_{bestask}}\sum_{n=1}^{\infty}[(z^{q_i+n-U} - z^{q_i-U})f_i(q,n) + (z^{|q_i-n|-U} - z^{q_i-U})g_i(q,n)]\\
&+ \sum_{i\in(i_{bestbid},i_{bestask})}\sum_{n=1}^{\infty}[(z^{n-U} - z^{-U})f_i(q,n) + (z^{n-U} - z^{-U})g_i(q,n)].
\end{aligned}$$

Then by Assumption 22,

$$\begin{aligned}
\tilde{\mathscr{Q}}V(q) &= \sum_{i\leq i_{bestbid}}\sum_{n=1}^{\infty}[(z^{-q_i+n-U} - z^{-q_i-U})g_i(q,n) + (z^{-q_i-n-U} - z^{-q_i-U})f_i(q,n)]\\
&+ \sum_{i\geq i_{bestask}}\sum_{n=1}^{\infty}[(z^{q_i+n-U} - z^{q_i-U})f_i(q,n) + (z^{q_i-n-U} - z^{q_i-U})g_i(q,n)]\\
&+ \sum_{i\in(i_{bestbid},i_{bestask})}\sum_{n=1}^{\infty}[(z^{n-U} - z^{-U})f_i(q,n) + (z^{n-U} - z^{-U})g_i(q,n)]\\
&= \sum_{i\leq i_{bestbid}} z^{-q_i-U}\sum_{n=1}^{\infty}(z^n - 1)[g_i(q,n) - \frac{f_i(q,n)}{z^n}]\\
&+ \sum_{i\geq i_{bestask}} z^{q_i-U}\sum_{n=1}^{\infty}(z^n - 1)[f_i(q,n) - \frac{g_i(q,n)}{z^n}]\\
&+ \sum_{i\in(i_{bestbid},i_{bestask})} z^{-U}\sum_{n=1}^{\infty}(z^n - 1)[f_i(q,n) + g_i(q,n)].
\end{aligned}$$

Using the definition of $G^{f,i,q}(z)$, $G^{g,i,q}(z)$, $f_i^*(q)$ and $g_i^*(q)$, we get

$$\begin{aligned}
\tilde{\mathscr{Q}}V(q) &= \sum_{i\leq i_{bestbid}} z^{-q_i-U}[g_i^*(q)(G^{g,i,q}(z) - 1) - f_i^*(q)(1 - G^{f,i,q}(z^{-1}))]\\
&+ \sum_{i\geq i_{bestask}} z^{q_i-U}[f_i^*(q)(G^{f,i,q}(z) - 1) - g_i^*(q)(1 - G^{g,i,q}(z^{-1}))]\\
&+ \sum_{i\in(i_{bestbid},i_{bestask})} z^{-U}[f_i^*(q)(G^{f,i,q}(z) - 1) + g_i^*(q)(G^{g,i,q}(z) - 1)].
\end{aligned}$$

Moreover, since for $z > 1$ we have $G^{f/g,i,q}(z^{-1}) < 1$ and $G^{f/g,i,q}(z) > 1$, we obtain

$$\begin{aligned}
\tilde{\mathcal{Q}}V(q) \leq & \sum_{i:i\leq i_{bestbid}, q_i<-U} z^{-q_i-U}(G^{g,i,q}(z)-1)\big(g_i^*(q)-f_i^*(q)\frac{1-G^{f,i,q}(z^{-1})}{G^{g,i,q}(z)-1}\big) \\
& + \sum_{i:i\geq i_{bestask}, q_i>U} z^{q_i-U}(G^{f,i,q}(z)-1)\big(f_i^*(q)-g_i^*(q)\frac{1-G^{g,i,q}(z^{-1})}{G^{f,i,q}(z)-1}\big) \\
& + \sum_{i:i\leq i_{bestbid}, q_i\geq -U} g_i^*(q)G^{g,i,q}(z) + \sum_{i:i\geq i_{bestask}, q_i\leq U} f_i^*(q)G^{f,i,q}(z) \\
& + \sum_{i\in(i_{bestbid}, i_{bestask})} [f_i^*(q)G^{f,i,q}(z) + g_i^*(q)G^{g,i,q}(z)] \\
\leq & \sum_{i:i\leq i_{bestbid}, q_i<-U} z^{-q_i-U}(G^{g,i,q}(z)-1)\big(g_i^*(q)-f_i^*(q)\frac{1-G^{f,i,q}(z^{-1})}{G^{g,i,q}(z)-1}\big) \\
& + \sum_{i:i\geq i_{bestask}, q_i>U} z^{q_i-U}(G^{f,i,q}(z)-1)\big(f_i^*(q)-g_i^*(q)\frac{1-G^{g,i,q}(z^{-1})}{G^{f,i,q}(z)-1}\big) \\
& + \sum_i [f_i^*(q)G^{f,i,q}(z)\mathbf{1}_{i>i_{bestbid}} + g_i^*(q)G^{g,i,q}(z)\mathbf{1}_{i<i_{bestask}}].
\end{aligned}$$

Now note that

$$\sup_q [f_i^*(q)G^{f,i,q}(z)\mathbf{1}_{i>i_{bestbid}} + g_i^*(q)G^{g,i,q}(z)\mathbf{1}_{i<i_{bestask}}]$$

is an increasing function of z. Thus by Assumption 23, we can find $z' > 1$ such that for any $z \leq z'$, $q \in \Omega$ and $i \in [-K, ..., K]$:

$$\big(f_i^*(q)G^{f,i,q}(z)\mathbf{1}_{i>i_{bestbid}} + g_i^*(q)G^{g,i,q}(z)\mathbf{1}_{i<i_{bestask}}\big) < L.$$

From Assumption 24, we see that we can find some $\tilde{z}$ with $1 < \tilde{z} \leq z'$ such that for any $(q,i), q_i < -U, i \leq i_{bestbid}$, for any $1 < z \leq \tilde{z}$:

$$g_i^*(q) - f_i^*(q)\frac{1-G^{f,i,q}(z^{-1})}{G^{g,i,q}(z)-1} < -r, \tag{II.8}$$

and for any $(q,i), q_i > U, i \geq i_{bestask}$, for any $1 < z \leq \tilde{z}$:

$$f_i^*(q) - g_i^*(q)\frac{1-G^{g,i,q}(z^{-1})}{G^{f,i,q}(z)-1} < -r. \tag{II.9}$$

Thus taking z in the definition of the function V satisfying $2Kz^{-U} > 1$ (so that the function V is coercive) and $1 < z \leq \tilde{z}$, we obtain

$$\begin{aligned}
\tilde{\mathcal{Q}}V(q) \leq & -r \sum_{i:i\leq i_{bestbid}, q_i<-U} z^{-q_i-U}(G^{g,i,q}(z)-1) \\
& -r \sum_{i:i\geq i_{bestask}, q_i>U} z^{q_i-U}(G^{f,i,q}(z)-1) + 2KL. \\
\leq & -rB_g(z) \sum_{i:i\leq i_{bestbid}, q_i<-U} z^{-q_i-U} \\
& -rB_f(z) \sum_{i:i\geq i_{bestask}, q_i>U} z^{q_i-U} + 2KL.
\end{aligned}$$

By Assumption 25, $B := \min(B_g(z), B_f(z)) > 0$. Therefore we get

$$
\begin{aligned}
\tilde{\mathscr{Q}}V(q) &\leq -rB \sum_{i:|q_i|>U} z^{|q_i|-U} + 2KL \\
&\leq -rB \sum_i z^{|q_i|-U} + 2K(L+rB) \\
&= -rBV(q) + 2K(L+rB).
\end{aligned}
$$

Finally, remark that in our setting, any compact set included in Ω is finite. A singleton being a petite set and a finite union of singletons remaining a petite set, see Proposition 5.5.5 in Meyn and Tweedie (2009), we get that all the compact sets are petite. Therefore by Theorem 6.1 in Meyn and Tweedie (1993), $q(t)$ is non-explosive and V-uniformly ergodic. Furthermore, by Theorem 4.2 in Meyn and Tweedie (1993) it is positive Harris recurrent.

7.2 Proof of Theorem 2

For some $1 < z \leq z^*$, set again

$$
V(q) = \sum_{i=-K, i\neq 0}^{K} z^{|q_i|-U}.
$$

We write $\tilde{\mathscr{P}}_{q,q'}$ the transition probability from q to q'. In the same way as in the preceding proof, we have

$$
\begin{aligned}
\Delta V(q) &:= \sum_{q'\in\Omega} \tilde{\mathscr{P}}_{q,q'}(V(q') - V(q)) \\
&= \sum_{i\leq i_{bestbid}} z^{-q_i-U}[b_i^*(q)(G^{g,i,q}(z)-1) - a_i^*(q)(1-G^{f,i,q}(z^{-1}))] \\
&\quad + \sum_{i\geq i_{bestask}} z^{q_i-U}[a_i^*(q)(G^{f,i,q}(z)-1) - b_i^*(q)(1-G^{g,i,q}(z^{-1}))] \\
&\quad + \sum_{i\in(i_{bestbid}, i_{bestask})} z^{-U}[a_i^*(q)(G^{f,i,q}(z)-1) + b_i^*(q)(G^{g,i,q}(z)-1)].
\end{aligned}
$$

Following the same method as in the proof of Theorem 1, we can easily find $1 < z' \leq z^*$ and $B > 0$ such that taking $z = z'$ in the definition of V, it is coercive and we get

$$
\Delta V(q) \quad \leq \quad -rV(q) + B.
$$

Now define the set $C := \{q,\ rV(q) \leq 2B\}$, C is obviously a finite set and is therefore petite. Furthermore, we have

$$
\Delta V(q) \leq -\frac{r}{2} V(q) + B\mathbf{1}_{q\in C}.
$$

Thus by Theorem 16.1.2 in Meyn and Tweedie (2009), $q(n)$ is V-uniformly ergodic.

Eventually, the fact that the chain is positive Harris recurrent is deduced from Theorem 9.1.8 together with Theorem 15.0.1 in Meyn and Tweedie (2009).

7.3 Proof of Theorem 3

For ease of notation, we write V instead of V^z. Let Q^{inc}, Q^{dec}, Q_K, Q_{-K} be four random variables such that $Q^{inc} \sim \pi^{inc}$, $Q^{dec} \sim \pi^{dec}$, $Q_K \sim \pi_K$ and $Q_{-K} \sim \pi_{-K}$. We define

$$\begin{aligned}
u^*(q) &= \frac{u(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\
d^*(q) &= \frac{d(q)}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\
n^*(q) &= \frac{\sum_i [f_i^*(q) + g_i^*(q)]}{\sum_i [f_i^*(q) + g_i^*(q)] + u(q) + d(q)} \\
\mathbb{E}_K &= \mathbb{E}[z^{Q_K - U}] \\
\mathbb{E}_{-K} &= \mathbb{E}[z^{Q_{-K} - U}] \\
\mathbb{E}_{\pi^{inc}} &= \mathbb{E}[V([Q^{inc}, c])] \\
\mathbb{E}_{\pi^{dec}} &= \mathbb{E}[V([Q^{dec}, c])].
\end{aligned}$$

Remarking that $V([q,c])$ does not depend on c, we write from now on $V(q)$ instead of $V([q,c])$. Moreover, we set $\mathscr{P}_{[q,c],[q',c']}$ as the transition probability from state $[q,c]$ to state $[q',c']$ and $\tilde{\mathscr{P}}_{q,q'}$ as the transition matrix of the embedded chain $q(n)$ when $u = d = 0$. Using the form of the infinitesimal generator $\mathscr{Q}$, we deduce

$$\begin{aligned}
\Delta V([q,c]) &:= \sum_{(q',c') \in \Omega \times \{-\alpha, \alpha\}} \mathscr{P}_{[q,c],[q',c']} (V(q') - V(q)) \\
&= n^*(q) \sum_{q' \in \Omega} \tilde{\mathscr{P}}_{q,q'} (V(q') - V(q)) \\
&\quad + u^*(q) \big[(1 - \theta^{reinit})(\mathbb{E}_K - z^{|q_{-K}| - U}) + \theta^{reinit} (\mathbb{E}_{\pi^{inc}} - V(q)) \big] \\
&\quad + d^*(q) \big[(1 - \theta^{reinit})(\mathbb{E}_{-K} - z^{|q_K| - U}) + \theta^{reinit} (\mathbb{E}_{\pi^{dec}} - V(q)) \big].
\end{aligned}$$

By Assumption 27, we have

$$\Delta V([q,c]) \leq n^*(q) \sum_{q' \in \Omega} \tilde{\mathscr{P}}_{q,q'} (V(q') - V(q)) + 2L^K + 2L^\pi.$$

Then as in the proof of Theorem 2, we can find $1 < z' \leq z^*$ and $B > 0$ such that taking $z = z'$ in the definition of V, it is coercive and we get

$$\Delta V(q) \leq -n^*(q) r V(q) + B n^*(q) + 2L^K + 2L^\pi.$$

Moreover, Assumption 28 ensures that for any q except those belonging to the finite set W, $n^*(q) > M$ with $M \in (0,1]$. Consequently,

$$\Delta V(q) \leq -MrV(q) + B + 2L^K + 2L^\pi + \nu \mathbf{1}_{[q,c] \in W \times \{-\alpha, \alpha\}},$$

with $\nu > 0$. Now define the set $C := \{[q,c],\ MrV(q) \leq 2B + 4L^K + 4L^\pi\}$. Being finite, $C \cup W$ is a petite set and we have

$$\Delta V([q,c]) \leq -\frac{Mr}{2} V([q,c]) + (B + 2L^K + 2L^\pi + \nu) \mathbf{1}_{[q,c] \in C \cup W}.$$

Hence from Theorem 16.1.2 in Meyn and Tweedie (2009), $Y(n)$ is V-uniformly ergodic.

Eventually, the fact that the chain is positive Harris recurrent is deduced from Theorem 9.1.8 together with Theorem 15.0.1 in Meyn and Tweedie (2009).

7.4 Proof of Theorem 5

7.4.1 Preliminary lemma

We start with the following preliminary lemma.

Lemma 1. *For the Markov chain $(q(n), c(n), \tau(n))$, the Cartesian product of any finite set included in $\Omega \times \{-\alpha, \alpha\}$ and $\mathbb{R}^+$ is petite.*

Proof. We first show that for any $q \in \Omega$ and $c \in \{-\alpha, \alpha\}$, the set $q \times c \times \mathbb{R}^+$ is petite (actually small). We define the measure $\nu_{q,c,\tau}$ which is so that for any $B \in \mathscr{B}(\Omega \times \{-\alpha, \alpha\} \times \mathbb{R}^+)$, $\nu_{q,c,\tau}(B)$ is the transition probability from $[q, c, \tau]$ to B in a single step:

$$\nu_{q,c,\tau}(B) = \mathscr{P}_{[q,c,\tau],B}.$$

Recall that in our framework, the transition probabilities from $(q(n), c(n), \tau(n))$ depend only on the value of $q(n)$. So we can write $\nu_{q,c,\tau}(B)$ as $\nu_q(B)$. In the sense of Equation (5.43) in Meyn and Tweedie (2009), the transition probability $\mathscr{P}_{[q,r,\tau],B}$ can be seen as a sampling kernel for the Markov chain $(q(n), c(n), \tau(n))$, using the Dirac measure at point 1 on $\mathbb{Z}^+$ as sampling measure. Moreover, for any $\tau \in \mathbb{R}^+$ and any $B \in \mathscr{B}(\Omega \times \{-\alpha, \alpha\} \times \mathbb{R}^+)$, we have

$$\mathscr{P}_{[q,c,\tau],B} \geq \nu_q(B).$$

Therefore the set $q \times c \times \mathbb{R}^+$ is petite. Then, as the union of two petite sets remains petite, see Proposition 5.5.5 in Meyn and Tweedie (2009), we have the result. □

7.4.2 A law of large numbers

In the next proposition, we give a law of large numbers for the inter-arrival times, which is the key element to establish the diffusive behavior of the price in calendar time. Within the proof of this proposition, we show that $(q(n), c(n), \tau(n))$ is positive Harris recurrent.

Proposition 1. *Let τ_i be the inter-arrival time between the i-th and the $i-1$-th jumps of the Markov process X. Under Assumptions 22, 23, 25, 26, 27, 28 and 29, almost surely, we have*

$$\frac{1}{n}\sum_{i=1}^{n} \tau_i \underset{n\to\infty}{\to} \mathbb{E}_{\pi^{**}}[\tau(1)],$$

with π^{**} the invariant distribution of $(q(n), c(n), \tau(n))$.

Proof. First we show that the Markov chain $(q(n), c(n), \tau(n))$ is positive Harris recurrent. For the Markov chain $(q(n), c(n))$, we have already proved that a coercive function V can be found, such that the following drift condition is satisfied for some $a > 0$ and $L < \infty$:

$$\Delta V([q, c]) \leq -aV([q, c]) + L.$$

Now, for the Markov chain $(q(n), c(n), \tau(n))$, take

$$V^*([q, c, t]) = V([q, c]) + t.$$

With obvious notation, we have

$$\begin{aligned} \Delta V^*([q, c, t]) &= \Delta V([q, c]) + \mathbb{E}_{[q,c,t]}[\tau] - t \\ &\leq -aV([q, c]) + L + \mathbb{E}_x[\tau] - t. \end{aligned}$$

Taking $a' = min(a,1)$ and $L' = L + 1/m$, using Assumption 29, we get

$$\begin{array}{rcl}\Delta V^*([q,c,t]) & \leq & -aV([q,c]) + L + 1/m - t \\ & \leq & -a'[V([q,c]) + t] + L + 1/m \\ & \leq & -a'V^*([q,c,t]) + L'.\end{array}$$

Now let $C = \{([q,c,t]),\ a'V^*([q,c,t]) \leq 2L'\}$. We have

$$\Delta V^*([q,c,t]) \quad \leq \quad -\frac{a'}{2}V^*([q,c,t]) + L'\mathbf{1}_{[q,c,t]\in C}.$$

According to Lemma 1, the set C is petite. Thus we can apply Theorem 15.0.1 in Meyn and Tweedie (2009) to deduce that the Markov chain $(q(n), r(n), \tau(n))$ is positive recurrent and thus admits an invariant measure.

Now remark that the function V^* is unbounded off petite sets (using Lemma 1 together with the fact that any subset of a petite set is itself petite). Consequently, Theorem 9.1.8 in Meyn and Tweedie (2009) enables us to obtain that the Markov chain $(q(n), c(n), \tau(n))$ is Harris recurrent. Therefore it is positive Harris recurrent.

We end the proof thanks to Theorem 17.0.1 from Meyn and Tweedie (2009). □

7.4.3 End of the proof of Theorem 5

We have

$$\tilde{S}^{(n)}(t) = \frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)]\rfloor)}{\sqrt{n}} + \Big(\frac{Z(N(nt))}{\sqrt{n}} - \frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)]\rfloor)}{\sqrt{n}}\Big).$$

According to Proposition 1, the sequence of processes $N(nt)/n$ converges to $t/\mathbb{E}_{\pi^{**}}[\tau(1)]$. Moreover, the limit of $Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)]\rfloor)/\sqrt{n}$ is continuous. Thus, using Skorohod representation theorem together with continuity properties in Skorohod topology, see Proposition VI.2.1 in Jacod and Shiryaev (2013), we get that the second term on the right hand side of the above equality tends to zero. Finally, from Theorem 4, we get

$$\frac{Z(\lfloor tn/\mathbb{E}_{\pi^{**}}[\tau(1)]\rfloor)}{\sqrt{n}} \overset{n\to\infty}{\to} \frac{\sigma}{\sqrt{\mathbb{E}_{\pi^{**}}[\tau(1)]}}B_t.$$

Combining these two results, we obtain the weak convergence of the rescaled price process to a Brownian motion with variance $\sigma^2/\mathbb{E}_{\pi^{**}}[\tau(1)]$, which concludes the proof.

Part II

Tick Size Effects

CHAPTER III

The Effect of Tick Value Changes on Market Microstructure: Analysis of the Japanese Experiements 2014

Abstract

The tick value is a crucial component of market design and is often considered the most suitable tool to mitigate the effects of high frequency trading. The goal of this paper is to demonstrate that the approach introduced in Dayri and Rosenbaum (2012) allows for an ex ante assessment of the consequences of a tick value change on the microstructure of an asset. To that purpose, we analyze the pilot program on tick value modifications started in 2014 by the Tokyo Stock Exchange in light of this methodology. We focus on forecasting the future cost of market and limit orders after a tick value change and show that our predictions are very accurate. Furthermore, for each asset involved in the pilot program, we are able to define (ex ante) an optimal tick value. This enables us to classify the stocks according to the relevance of their tick value, before and after its modification.

1 Introduction

On January 14, 2014, the Tokyo Stock Exchange (TSE) launched the first phase of its pilot program on tick value[1] modifications, reducing the tick value of the TOPIX 100 index stocks priced above ¥3000 by approximately 90% (see Section 3.1 for more details on this pilot program). The second phase was implemented on July 22, 2014, targeting a sub-Yen tick value reduction for stocks of the TOPIX 100 index priced below ¥5000. The third phase of this program is expected to start in September 2015, when a new tick value table should be announced after the evaluation of the effects of the tick value reductions in the first two phases.

The tick value is probably the most relevant tool that can be used by exchanges and regulators in order to improve the trading quality and the robustness of the market structure, see Lehalle et al. (2014). Compared with other more controversial proposals, such as imposing a minimum resting time for orders to remain valid or using frequent batch auctions, setting a suitable tick value is in general considered to be a better way to control the growing activity of high frequency traders, which accounts nowadays for more than 40% of the total volume on equity markets. Indeed, a tick value change induces very little cost and is easily reversible if the outcome does not meet the market designer's expectations.

[1]The tick value is the minimum price variation allowed for an asset on a given market.

The Tokyo Stock Exchange is not alone in its search for better tick values. In the United States, on August 26, 2014, the Securities and Exchange Commission announced a program aiming at widening tick values for stocks with small capitalization. A targeted 12 months pilot experiment will be implemented to assess the effects of such changes. In Europe, in May 2014, the European Securities and Markets Authority released a MiFID2/R discussion paper which proposes two options for a new harmonized tick value regime to be introduced across all trading venues. These two options are currently debated by European regulators.

Before the pilot program, most Japanese stocks were typical examples of *large tick assets*, that is assets whose bid-ask spread is almost always equal to one tick. The tick value being the lower bound for the bid-ask spread, when it is too large, the cost of market orders becomes very significant. This not only damages liquidity takers but also the "slow" liquidity providers who suffer from the intensification of the speed competition for gaining time priority in the order book queues, see Moallemi and Yuan (2015). Although it is quite commonly accepted that it is preferable to reduce the tick value for these large tick assets, finding the appropriate tick value remains a very intricate problem. Indeed, most of the numerous works about the consequences of a tick value change are empirical and focus on the outcomes of this market design modification in an *ex post* basis, see for example Lau and McInish (1995), Ahn, Cao, and Choe (1996), Bacidore (1997), Bessembinder (2000), Goldstein and Kavajecz (2000), Chung and Van Ness (2001), Chung and Chuwonganant (2002), Bourghelle and Declerck (2004) and Wu, Krehbiel, and Brorsen (2011).

These studies have clearly shown that a change in the tick value may lead to significant implications for the bid-ask spread, the available volume in the order book and many other microstructural quantities. However, very few quantitative tools exist for predicting *ex ante* these effects. Consequently, exchanges and market regulators often rely on the trial and error approach in order to set appropriate tick values. The Japanese pilot experiment, in which the tick value reduction program is conducted in three phases, is one of such examples. Indeed, the effects of tick value changes in the first two phases are evaluated ex post to help the design of a new tick value table to be implemented in the last phase.

In Dayri and Rosenbaum (2012), the authors build a quantitative approach towards solving the crucial problems of forecasting the consequences of a tick value change and determining an optimal tick value. To that purpose, based on the model with uncertainty zones introduced in Robert and Rosenbaum (2011), they use the key microstructural indicator η (half the ratio between price continuations and alternations, see Section 2) which summarizes the high frequency features of an asset. The paramount importance of the parameter η is due to the fact that there is a one to one bijection between its value and the cost of market and limit orders. We recall this connection in details in Section 2. Hence, measuring η allows us to classify stocks according to whether they are profitable for market makers or rather balanced. Furthermore, being able to predict the consequences of a tick value change on η means one can anticipate the new microstructural costs induced by this tick value modification, which is precisely what exchanges and regulation authorities need. Such predictions are possible using the approach in Dayri and Rosenbaum (2012) where explicit forecasting formulas for the parameter η are provided. Moreover, the way to set a tick value leading to an optimal η is also established (see Section 2 for our definition of optimality).

In this work, our goal is to show that the theoretical forecasting formulas in Dayri and Rosenbaum (2012) do enable us to predict ex ante the consequences of a tick value change on the microstructure of an asset, notably on the trading costs. To demonstrate this, we use 18 months of tick by tick market data from the TSE, including the whole year 2014 when the pilot program is in place. Very accurate results are obtained for the prediction of the parameter η. Thus, the approach in Dayri and Rosenbaum (2012) is proved to be indeed very helpful in both predicting the consequences of a tick value change and choosing an optimal tick value for large tick assets.

The paper is organized as follows. We recall in Section 2 the reading of the model with uncertainty zones as a mean to quantify the average cost of market and limit orders using the crucial microstructural indicator η. At the end of this section, we give the prediction formula for η after a tick value modification. Hence we provide a way to predict the change in the cost of market and limit orders induced by such modification. In Section 3, we consider the TSE pilot experiment on tick values. First, before the start of the program, we classify Japanese assets in two categories: stocks with costly market orders and stocks with balanced costs between market and limit orders. Note that the situation of costly limit orders is very unlikely. Indeed, in that case, market makers would increase their spread, what they can always do. Then we apply in the same section the forecasting methodology presented in Section 2. In particular, we predict whether a stock will change category after the tick value modification. We conclude in Section 4.

2 Cost of trading and high frequency price dynamics

2.1 The model with uncertainty zones: When the tick prevents price discovery

The model with uncertainty zones, introduced in Robert and Rosenbaum (2011), is a high frequency model for the transaction prices of a large tick asset. It reproduces most macroscopic and microscopic stylized facts of price dynamics and is very suitable for the analysis of the role of the tick value in determining the microstructural features of an asset. This model assumes the existence of a latent efficient price X_t, typically a martingale, and states that a transaction can occur at a given price level (on the tick grid) only provided this price level is close enough to the efficient price. This proximity is quantified by the parameter η: the distance between the potential transaction price and the efficient price has to be smaller than $\alpha/2+\eta\alpha$, with α the tick value of the asset. Thus, for a large tick asset, assuming the efficient price lies within the one tick bid-ask spread $[b, b+\alpha]$, we have $\eta \in [0, 1/2]$ and obtain three zones for the efficient price:

- If it lies between b and $b+\alpha(1/2-\eta)$, transactions can only occur on the bid side (bid zone).
- If it lies between $b+\alpha(1/2+\eta)$ and $b+\alpha$, transactions can only occur on the ask side (ask zone).
- If it lies between $b+\alpha(1/2-\eta)$ and $b+\alpha(1/2+\eta)$, transactions can occur both on the bid and on the ask side (buy/sell or uncertainty zone).

These three zones are summarized in Figure III.1.

Estimation of η The parameter η can be very easily estimated as follows. We define an alternation (resp. continuation) as a transaction price jump of one tick whose direction is opposite to (resp. the same as) the one of the preceding transaction price jump. Let $N^{(a)}$ and $N^{(c)}$ be respectively the number of alternations and continuations during the period $[0, t]$. The estimator of η over $[0, t]$ is simply given by

$$\widehat{\eta} = \frac{N^{(c)}}{2N^{(a)}}.$$

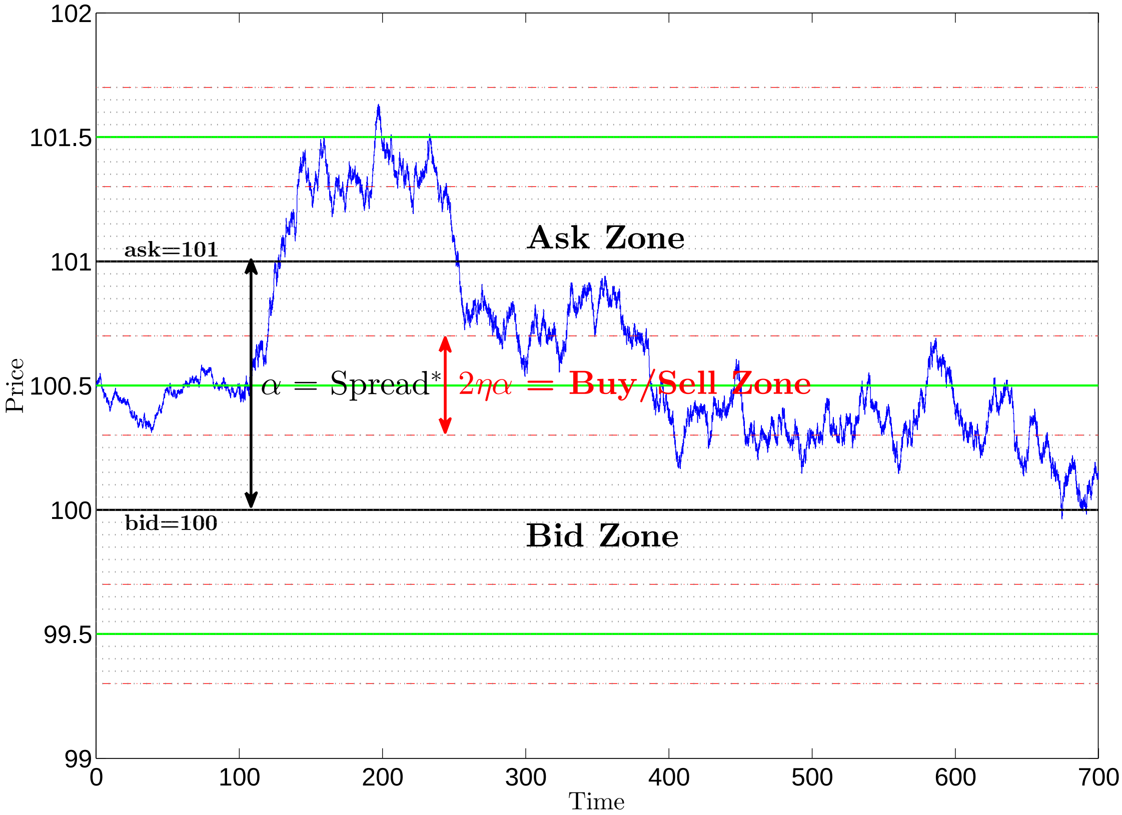

Figure III.1: The three different zones when the bid-ask is 100-101 and the tick value is equal to one. The red dotted lines are the limits of the uncertainty zone. The uncertainty zone inside the spread is the buy/sell zone. The upper dotted area is the ask zone and the lower dotted area is the bid zone.

Theoretical properties for this estimator are established in Robert and Rosenbaum (2012). Note that in Section 3, the estimated values of η over given time periods of several months will be given by the averages of the daily estimations of η over all the days of the periods.

2.2 Perceived tick size and cost of market orders

The parameter η controls the width of the uncertainty zones (which is $2\eta\alpha$; when the efficient price is inside this zone, investors cannot clearly decide if it is more relevant to buy or sell) and measures the bouncing intensity of the transaction price due to the existence of the tick value. It can actually be seen as an indicator for the perceived tick size of a large tick asset: A very small η ($\eta \ll 0.5$) means that for market participants, the tick value appears much too large (in such case, it is necessary to be sharp in term of estimation of the efficient price to know if it is reasonable to buy or sell at a given time), while a η close to $1/2$ is synonym of a suitable tick value (in such situation, the uncertainty zones almost correspond to the tick grid). To understand this, consider a market order of unit volume at price P_t at time t. Its cost with respect to the efficient price is $P_t - X_t$. For a large tick asset, the average cost of such market order can be computed and is equal to

$$\alpha/2 - \eta\alpha,$$

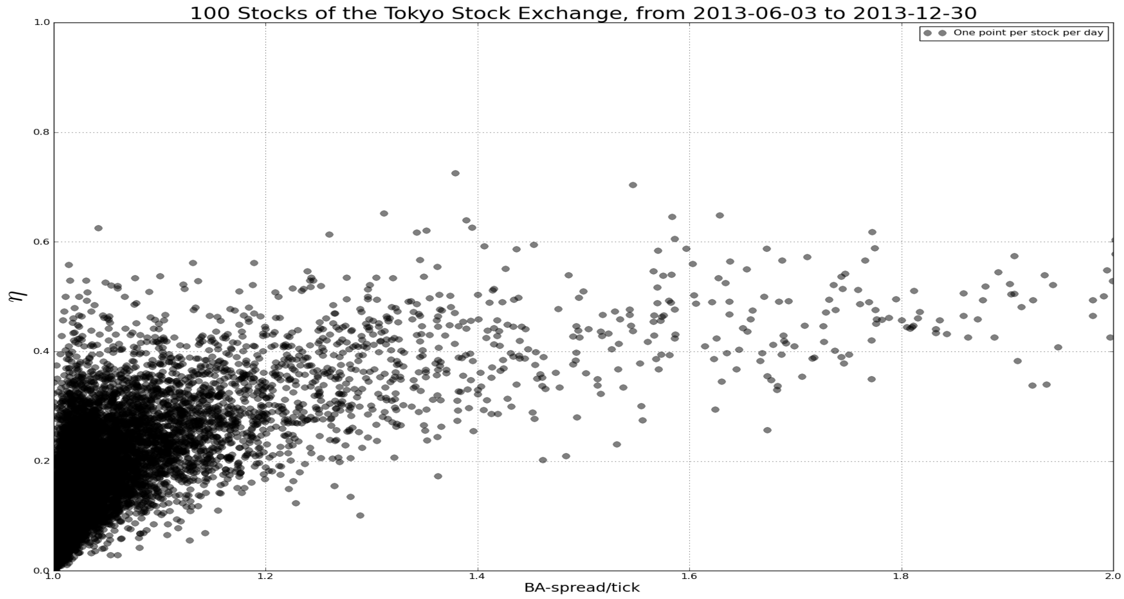

Figure III.2: Daily values of η for all the assets of the TOPIX 100 index, for every day so that the daily average spread is smaller than two ticks, from 2013, June 3 to 2013, December 30.

see Dayri and Rosenbaum (2012). The quantity $\alpha/2 - \eta\alpha$ is non negative provided $\eta \le 0.5$, a condition which is almost systematically satisfied by estimated values of η on large tick assets, see Figure III.2. Indeed, it would otherwise mean that on average, market makers lose money. This is not really possible since in such situation, they would simply increase their spread, what they can always do. Thus, when $\eta < 0.5$, market takers have to pay a fixed positive cost in order to get liquidity, while market makers gain profit by placing limit orders[2]. This cost paid by liquidity takers is given by $\alpha/2 - \eta\alpha$. Consequently, for large tick assets, the classical efficiency rule that assumes zero cost for both market and limit orders becomes irrelevant.

2.3 Implicit bid-ask spread and cost of limit orders

Within bid-ask quotes of the form $[b, b+\alpha]$, the width of the uncertainty zone represents the range of values for the efficient price X_t where transactions can occur both at the best bid and the best ask side. The size of this range is $2\eta\alpha$. Therefore, it is natural to view the quantity $2\eta\alpha$ as an ***implicit spread***. This idea is fully supported by the regression analysis in Dayri and Rosenbaum (2012), which shows empirically that for large tick assets, $\eta\alpha$ is proportional to the volatility per trade:

$$\eta\alpha \sim c\frac{\sigma}{\sqrt{M}}, \tag{III.1}$$

where σ denotes the square root of the daily integrated variance of the price, M the number of transactions per day and c a constant around one. Yet it is well-known that for small tick assets, for which the (conventional) spread can evolve freely and is not artificially bounded from below by the tick value, the average spread is proportional to the volatility per trade, see Madhavan, Richardson, and Roomans (1997) and Wyart, Bouchaud, Kockelkoren, Potters, and Vettorazzo

[2]Of course this is true only considering the aggregated group of market makers. In practice, the gains of individual market makers are often small since there are several of them and the queues at the best bid and ask levels are quite long.

(2008). This confirms that $2\eta\alpha$ can be interpreted as an implicit spread for a large tick asset.

Actually the fact that for small tick assets, the average spread S is proportional to the volatility per trade simply comes from the efficiency condition stating that market makers make on average zero profit due to competition. More precisely, to derive the spread-volatility per trade relation, let us consider a dichotomy between market makers using limit orders and market takers using market orders. In that case, the average profit and loss per trade of a typical market making strategy, which can be understood as that of a limit order, is essentially equal to $S/2 - c\sigma/\sqrt{M}$, see Wyart, Bouchaud, Kockelkoren, Potters, and Vettorazzo (2008). Therefore, the efficiency assumption implies

$$\frac{S}{2} \sim c\frac{\sigma}{\sqrt{M}}.$$

In the case of a large tick asset, for which $S = \alpha$, as seen in the previous subsection, market orders are costly, their cost being on average $\alpha/2 - \eta\alpha$. Therefore the profit and loss of market makers, which is still $S/2 - c\sigma/\sqrt{M}$, is no longer zero. Indeed, it is precisely the cost paid by market takers. Consequently, we get

$$S/2 - c\frac{\sigma}{\sqrt{M}} = \alpha/2 - c\frac{\sigma}{\sqrt{M}} = \alpha/2 - \eta\alpha,$$

which leads to Equation (III.1). Importantly, this simple cost analysis and Equation (III.1) derived from it enable us to design simple prediction formulas for η after a change in the tick value.

2.4 Prediction of the cost of market and limit orders

Based on the fact that Equation (III.1) should hold for a large tick asset for any tick value, the authors in Dayri and Rosenbaum (2012) establish three prediction formulas for the new value of η, and therefore for the new cost of market and limit orders, after a change in the tick value of an asset. Each of the three formulas corresponds to different assumptions. For simplicity, we only present here a formula which does not require any prior regression analysis and assumes a linear shape for the cumulative latent liquidity.

Let us consider a large tick asset for which the current tick value is α_0 and associated is η_0. Then, if the tick value is changed to α, the formula for the new parameter η gives:

$$\eta \sim (\eta_0 + 0.1)(\frac{\alpha_0}{\alpha})^{1/2} - 0.1. \qquad \text{(III.2)}$$

We now comment this formula and its use to predict the new microstructural features of an asset after a tick value change:

- Formula (III.2) actually holds only provided the asset remains a large tick asset after the change in the tick value. However, due to the concurrence mechanism, market makers maintain a spread equal to one tick as long as they make profit from it, that is as long as $\eta < 1/2$. The value of η in the formula being decreasing with α, we get that Formula (III.2) holds provided $\alpha \geq \alpha^*$ with
$$\alpha^* = (\frac{\eta_0 + 0.1}{0.6})^2 \alpha_0.$$
- Formula (III.2) enables us to tell whether the asset remains a large tick asset after the tick value change: if the forecast value of η is greater than $1/2$ (that is $\alpha < \alpha^*$), the asset is predicted to become a small tick asset after the tick value modification. However, note that in that case, the forecast value of η cannot really be interpreted beyond this (becoming small tick or not).

- If the predicted value of η is smaller than $1/2$, Formula (III.2) provides the estimated η after the tick value change, and therefore the estimated cost of market and limit orders. In particular, this allows us to tell ex ante whether a stock will become/remain favorable for market makers or exhibit balanced trading costs. This is probably the most relevant viewpoint in term of regulation.

2.5 What is a suitable tick value?

From a regulatory perspective, a tick value can probably be seen as suitable if:

- The bid-ask spread is close to one tick, ensuring the presence of liquidity in the order book.
- Transaction costs are close to zero for market orders. In that case, the market is efficient and market makers do not take advantage of the tick value to the detriment of final investors acting mainly as liquidity takers.

Thus, in our approach, an asset enjoys a relevant tick value if it is a large tick asset and its η parameter is close to $1/2$. Indeed, recall that the cost of a market order for a large tick asset is $\alpha/2 - \eta\alpha$.

Note that according to Formula (III.2), starting from a large tick asset, the optimal tick value can be obtained setting $\alpha = \alpha^*$. With this optimality notion in mind, we conduct in the next section an empirical analysis of the Japanese pilot program on tick values.

3 Analysis of the Tokyo Stock Exchange pilot program on tick values

3.1 Data description

We use data from the 55 Japanese stocks of the TOPIX 100 index involved in the pilot program in 2014. Our database, provided by Capital Fund Management, records the time and price of every transaction, as well as the best bid and ask prices right before the transactions, from June 3, 2013 to December 30, 2014. We remove market data corresponding to the first and last hour of trading, as these periods have usually specific features due to the opening/closing auction. Three different phases are distinguished in this study:

- Phase 0 (before the pilot program): from June 3, 2013 to January 13, 2014.
- Phase 1 (from the first implementation of the tick value reduction program to the second one): from January 14, 2014 to July 21, 2014.
- Phase 2 (after the second implementation of the tick value reduction program): from July 22, 2014 to December 30, 2014.

The details of the pilot program are given in Table III.1.

Quoted price below (¥)	Phase 0 tick value (¥)	Phase 1 tick value (¥)	Phase 2 tick value (¥)
1,000	1	1	0.1
3,000	1	1	0.5
5,000	5	1	0.5
10,000	10	1	1
30,000	10	5	5
50,000	50	5	5
100,000	100	10	10
300,000	100	50	50
500,000	500	50	50
1,000,000	1000	100	100
3,000,000	1000	500	500
5,000,000	5000	500	500
10,000,000	10000	1000	1000
30,000,000	10000	5000	5000
50,000,000	50000	5000	5000
Higher prices	100000	10000	10000

Table III.1: Tick value reduction table.

3.2 Classification of the stocks in Phase 0

In this work, we use the average spread in number of ticks, now denoted by S and estimated as the average value of the bid-ask spread right before a transaction, to classify the stocks into three groups:

- Small tick stocks: $S > 1.6$.
- Large tick stocks: $S \leq 1.5$.
- Ambiguous case between large and small tick: $1.5 < S \leq 1.6$.

For large tick stocks, we use the parameter η to distinguish between balanced stocks, for which market orders are reasonably costly, and stocks where market makers (viewed again as an aggregated class) obtain significant profit from liquidity takers thanks to the tick value. We use the following criterion:

- Balanced stocks: $\eta \geq 0.4$.
- Market makers favorable stocks: $\eta < 0.4$.

Note that small tick stocks will be considered balanced stocks. However, they do not fulfill the criteria for stocks having a suitable tick value, their spread being somehow too large.

For each of the 55 stocks, for Phase 0, we give in Table III.2 the average spread (S_0), the value of the η parameter (η_0) and tell whether the stock is large tick or not (Yes or No for the variable LTick_0) and whether it is balanced or not (Yes or No for the variable Bal_0).

We see that all the assets but one (Mitsubishi Estate Co Ltd) are large tick stocks in Phase 0. However, among the remaining 54 large tick stocks, only five of them are balanced: Chubu Electric Power Co Inc, Inpex Corp, Kirin Holdings Co Ltd, Kubota Corp and Sumitomo Electric Industries Ltd. According to our framework, no tick value modification was necessary for these five assets. However, a tick value reduction can be beneficial for the 49 other stocks, which somehow justifies the will of the TSE to launch the pilot program.

Company Name	S_0	η_0	LTick_0 Large Tick	Bal_0 Balanced
Aeon Co Ltd	1.07	0.18	Yes	No
ANA Holdings Inc	1.23	0.09	Yes	No
Asahi Class Co Ltd	1.03	0.15	Yes	No
Asahi Kasei Corp	1.04	0.21	Yes	No
Astellas Pharma Inc	1.05	0.14	Yes	No
Bank of Yokohama Ltd	1.02	0.23	Yes	No
Canon Inc	1.04	0.06	Yes	No
Chubu Electric Power Co Inc	1.25	0.49	Yes	Yes
Daiwa Securities Group Inc	1.04	0.20	Yes	No
Dalichi Sankyo Co Ltd	1.15	0.32	Yes	No
Dai-ichi Life Insurance Co Ltd	1.10	0.22	Yes	No
Fujitsu Ltd	1.03	0.13	Yes	No
Hitachi Ltd	1.04	0.10	Yes	No
Honda Motor Co Ltd	1.04	0.10	Yes	No
Inpex Corp	1.27	0.55	Yes	Yes
ITOCHU Corp	1.05	0.18	Yes	No
Japan Tobacco Inc	1.04	0.12	Yes	No
JX Holdings Inc	1.01	0.08	Yes	No
Kansai Electric Power Co Inc	1.24	0.33	Yes	No
Kirin Holdings Co Ltd	1.43	0.53	Yes	Yes
Kubota Corp	1.34	0.46	Yes	Yes
Komatsu Ltd	1.30	0.29	Yes	No
Marubeni Corp	1.02	0.12	Yes	No
Mitsubishi Chemical Holdings Corp	1.01	0.16	Yes	No
Mitsubishi Corp	1.22	0.36	Yes	No
Mitsubishi Electric Corp	1.10	0.31	Yes	No
Mitsubishi Estate Co Ltd	1.70	0.60	No	Yes
Mitsubishi Heavy Industries Ltd	1.01	0.11	Yes	No
Mitsubishi UFJ Financial Group Inc	1.03	0.04	Yes	No
Mitsui Co Ltd	1.12	0.23	Yes	No
Mitsui Fudosan Co Ltd	1.07	0.23	Yes	No
Mizuho Financial Group Inc	1.14	0.07	Yes	No
Nissan Motor Co Ltd	1.06	0.14	Yes	No
Nippon Steel Sumitomo Metal Corp	1.02	0.05	Yes	No
Nippon Telegraph Telephone Corp	1.04	0.08	Yes	No
Nomura Holdings Inc	1.05	0.06	Yes	No
NTT DoCoMo Inc	1.28	0.24	Yes	No
ORIX Corp	1.21	0.33	Yes	No
Osaka Gas Co Ltd	1.04	0.16	Yes	No
Panasonic Corp	1.14	0.16	Yes	No
Resona Holdings Inc	1.00	0.07	Yes	No
Ricoh Co Ltd	1.13	0.36	Yes	No
Seven I Holdings Co Ltd	1.06	0.16	Yes	No
Softbank Corp	1.05	0.06	Yes	No
Sony Corp	1.17	0.24	Yes	No
Sumitomo Corp	1.04	0.18	Yes	No
Sumitomo Electric Industries Ltd	1.24	0.40	Yes	Yes
Sumitomo Mitsui Trust Holdings Inc	1.01	0.14	Yes	No
Sumitomo Mitsui Financial Group Inc	1.15	0.08	Yes	No
Takeda Pharmaceutical Co Ltd	1.06	0.13	Yes	No
Tokyo Gas Co Ltd	1.05	0.16	Yes	No
Tokio Marine Holdings Inc	1.05	0.18	Yes	No
Toray Industries Inc	1.03	0.13	Yes	No
Toshiba Corp	1.03	0.06	Yes	No
Toyota Motor Corp	1.03	0.04	Yes	No

Table III.2: Average spread, value of η and categories (large tick or not; balanced or not) for the 55 stocks in Phase 0.

3.3 Phase 0 - Phase 1

We now test the prediction formula (III.2) between Phase 0 and Phase 1. We first select 12 stocks among the 55 stocks based on the following criteria:

- These stocks are large tick assets during Phase 0.
- These stocks are involved in the tick value reduction program during Phase 1.
- For every stock, days on which multiple tick values are used[3] are removed from the database. We then choose the tick value right before the end of Phase 0 and the tick value right after the beginning of Phase 1 as two reference tick values. Stocks for which the numbers of days when the tick value is equal to its reference value in Phase 0 and Phase 1 are both greater than 10 are finally selected.

Twelve assets are remaining after this selection. For each of these stocks, based on the value of η in Phase 0 (η_0), we predict the new value of η in Phase 1 (η_1^p) using Formula (III.2). We provide confidence intervals based on the 25% and 75% quantiles of the distribution of the estimated daily η in Phase 0[4]. We also forecast whether the asset will be large tick in Phase 1 (LTick_1^p) and balanced in Phase 1 (Bal_1^p). More precisely, considering a predicted η larger than $1/2$ corresponds to an increase of the spread (recall that the situation $\eta > 1/2$ is not compatible with a one tick spread):

- If $\eta_1^p \geq 0.55$, the asset is predicted to become a small tick asset after the tick value change.
- If $\eta_1^p < 0.5$, the asset is predicted to remain a large tick asset after the tick value change, with the forecast value for the new η being meaningful and given by η_1^p.
- We qualify the situation $0.5 \leq \eta_1^p < 0.55$ as an "ambiguous" case between large tick and small tick.

We compare the predictions to the actual quantities in Phase 1: η_1, LTick_1 and Bal_1. The results are given in Table III.3.

The obtained average relative prediction error for η_1, that is the average of the $|\eta_1^p - \eta_1|/\eta_1$ is less than 18%. This shows that thanks to Formula (III.2), we can forecast the new value of η, and therefore the new trading costs, with a good accuracy. Our prediction for η being quite sharp, it is no surprise that we are able to forecast whether or not a stock is going to remain large tick and whether it is balanced in Phase 1. For nine of the stocks, our forecast was that it would remain large tick, and seven of these predictions were correct (classifying the "ambiguous" case as correct). We also predicted that two of the assets would become small tick and both predictions were correct. Thus, the predictive power of our methodology is very high.

According to our optimality notion, three of the twelve stocks now enjoy a suitable tick value (balanced large tick stocks): Seven I Holdings Co Ltd, Takeda Pharmaceutical Co Ltd and Tokio Marine Holdings Inc. Such results could have been obtained ex ante using our approach. Indeed, following our methodology, a regulator or an exchange can anticipate the suitable way to operate a tick value change, especially when the goal is to decrease the tick value of an unbalanced stock.

[3]We recall that the tick value of a stock depends on its price, see Table III.1.

[4]The estimated η being the average of the daily estimations, in very few cases, the prediction can fall out of the confidence interval. Then we replace the prediction by the closest bound in the confidence interval.

Company name	S_0	η_0	S_1	η_1	LTick$_1$	Bal$_1$	η_1^p	LTick$_1^p$	Bal$_1^p$
**Astellas Pharma Inc	1.05	0.14	1.72	0.43	No	Yes	0.66 [0.50,0.71]	No	Yes
**Canon Inc	1.04	0.06	1.13	0.23	Yes	No	0.26 [0.19,0.27]	Yes	No
**Honda Motor Co Ltd	1.04	0.10	1.23	0.32	Yes	No	0.34 [0.26,0.37]	Yes	No
**Japan Tobacco Inc	1.04	0.12	1.23	0.32	Yes	No	0.39 [0.26,0.41]	Yes	No
**Mitsui Fudosan Co Ltd	1.07	0.23	1.95	0.66	No	Yes	0.63 [0.52,0.69]	No	Yes
*Nippon Telegraph Telephone Corp	1.04	0.08	2.00	0.62	No	Yes	0.46 [0.35,0.51]	Yes	Yes
(*)*Seven I Holdings Co Ltd	1.06	0.16	1.55	0.51	Ambiguous	Yes	0.49 [0.38,0.55]	Yes	Yes
*Softbank Corp	1.05	0.06	1.85	0.50	No	Yes	0.40 [0.32,0.40]	Yes	Yes
*Sumitomo Mitsui Financial Group Inc	1.15	0.08	1.33	0.34	Yes	No	0.47 [0.27,0.47]	Yes	Yes
**Takeda Pharmaceutical Co Ltd	1.06	0.13	1.46	0.43	Yes	Yes	0.42 [0.28,0.45]	Yes	Yes
(*)*Tokio Marine Holdings Inc	1.05	0.18	1.39	0.46	Yes	Yes	0.53 [0.41,0.57]	Ambiguous	Yes
**Toyota Motor Corp	1.03	0.04	1.36	0.32	Yes	No	0.35 [0.27,0.33]	Yes	No

Table III.3: For the 12 selected stocks: Average spread and value of η in Phase 0 and Phase 1, categories in Phase 1, and predictions for η and the categories. The number of stars * in front of a company name represents the number of good predictions (one for being large tick or not, one for being balanced or not). A star between brackets (*) corresponds to an "ambiguous" case.

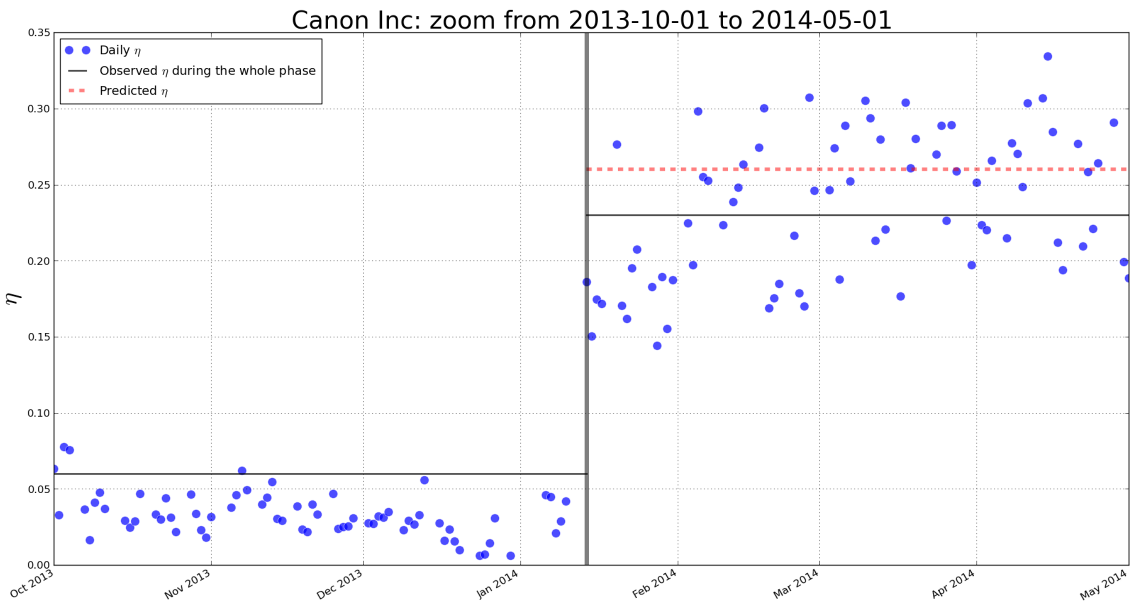

Figure III.3: Daily estimations of η for the last 3.5 months of Phase 0 and the first 3.5 months of Phase 1 for the stock Canon Inc.

To end this subsection, as an illustration, we give in Figure III.3 detailed results about η for the stock Canon Inc. More precisely, we provide daily estimations of η for the last 3.5 months of Phase 0 and the first 3.5 months of Phase 1. We also add the average values of η during both phases together with our forecast for the value of η in Phase 1. We see on this example that our prediction is very close to the realized value.

3.4 Phase 1 - Phase 2

The tick value reduction program affects much more stocks in Phase 2. Using the same selection criteria as previously (replacing Phase 0 by Phase 1 and Phase 1 by Phase 2), we find 48 stocks that are large tick assets during Phase 1 and have their tick value effectively reduced in Phase 2.

We draw a similar analysis as the one for Phase 0-Phase 1. The results are given in Table III.4, where the index 1 is used to denote quantities in Phase 1 and the index 2 quantities in Phase 2.

Once again, we obtain an excellent accuracy for predicting the value of η in Phase 2 based on that in Phase 1. Indeed, the average relative error is here less than 17%. Among the 48 assets, 16 of them are predicted to become small tick stocks and all these predictions are correct. Moreover, 28 stocks are predicted to remain large tick and 23 of these predictions are correct (taking the ambiguous cases as correct). Regarding the fact of being balanced, more than 85% of our predictions are realised. Hence the study of the evolution of the market between Phase 1 and Phase 2 confirms what was found for Phase 0-Phase 1: our device based on η enables us to forecast ex ante the consequences of a change in the tick value on the market microstructure.

Note that after this second phase, the stocks Canon Inc, Chubu Electric Power Co Inc, Dalichi Sankyo Co Ltd, JX Holdings Inc, Komatsu Ltd and Mitsubishi Electric Corp seem to have a suitable tick value (balanced large tick stocks).

4 Conclusion

Based on data from the TSE pilot program, we have studied the effects of tick value changes on the microstructure of large tick stocks. This has been done using the microstructural parameter η which summarizes the high frequency features of a large tick asset, in particular the associated trading costs. The prediction formula suggested in Dayri and Rosenbaum (2012) for the new value of the parameter η after a tick value change has been tested using all the stocks of the Japanese experiment. We have compared the prediction results in Phase 1 and Phase 2 with the realized η, and shown that Formula (III.2) provides very accurate forecasts. In particular, we can predict ex ante whether a large tick stock will become a small tick stock after a tick value change and whether or not its associated trading costs will be balanced between market makers and liquidity takers.

This work validates the quantitative tools developed in Dayri and Rosenbaum (2012) for studying the consequences of a tick value modification. It provides detailed practical guidelines for market regulators and exchanges searching for optimal tick values. Indeed, it can help them choose suitable tick values without applying any trial and error method, which may largely reduce the duration and cost of pilot programs.

Company Name	S_1	η_1	S_2	η_2	LTick_2	Bal_2	η_2^p	LTick_2^p	Bal_2^p
**Aeon Co Ltd	1.03	0.12	1.16	0.16	Yes	No	0.21 [0.18,0.26]	Yes	No
**Asahi Class Co Ltd	1.01	0.14	2.49	0.87	No	Yes	0.67 [0.52,0.82]	No	Yes
**Asahi Kasei Corp	1.03	0.23	3.43	1.02	No	Yes	0.93 [0.76,1.07]	No	Yes
**ANA Holdings Inc	1.00	0.02	1.37	0.30	Yes	No	0.29 [0.26,0.31]	Yes	No
**Bank of Yokohama Ltd	1.02	0.22	3.28	0.98	No	Yes	0.92 [0.77,1.05]	No	Yes
(*)Canon Inc	1.13	0.23	1.59	0.43	Ambiguous	Yes	0.36 [0.30,0.43]	Yes	No
**Chubu Electric Power Co Inc	1.10	0.31	1.45	0.44	Yes	Yes	0.48 [0.40,0.55]	Yes	Yes
**Daiwa Securities Group	1.02	0.19	3.42	1.09	No	Yes	0.81 [0.70,0.90]	No	Yes
**Dai-ichi Life Insurance Co Ltd	1.08	0.25	1.35	0.35	Yes	No	0.39 [0.33,0.47]	Yes	No
**Dalichi Sankyo Co Ltd	1.10	0.27	1.43	0.40	Yes	Yes	0.43 [0.35,0.50]	Yes	Yes
**Fujitsu Ltd	1.01	0.16	2.80	0.97	No	Yes	0.72 [0.60,0.80]	No	Yes
(*)*Hitachi Ltd	1.01	0.09	2.55	0.81	No	Yes	0.50 [0.42,0.58]	Ambiguous	Yes
*Honda Motor Co Ltd	1.23	0.32	1.69	0.49	No	Yes	0.49 [0.43,0.55]	Yes	Yes
*Inpex Corp	1.08	0.25	1.34	0.37	Yes	No	0.40 [0.35,0.45]	Yes	Yes
**ITOCHU Corp	1.03	0.13	1.17	0.24	Yes	No	0.23 [0.18,0.27]	Yes	No
(*)*Japan Tobacco Inc	1.23	0.32	1.86	0.55	No	Yes	0.50 [0.40,0.57]	Ambiguous	Yes
(*)*JX Holdings Inc	1.01	0.07	1.52	0.41	Ambiguous	Yes	0.44 [0.38,0.50]	Yes	Yes
**Kansai Electric Power Co Ltd	1.07	0.25	4.20	1.03	No	Yes	1.01 [0.86,1.17]	No	Yes
*Kirin Holdings Co Ltd	1.10	0.29	1.31	0.34	Yes	No	0.45 [0.31,0.57]	Yes	Yes
*Komatsu Ltd	1.10	0.24	1.50	0.46	Yes	Yes	0.39 [0.33,0.43]	Yes	No
**Kubota Corp	1.15	0.37	1.69	0.59	No	Yes	0.57 [0.49,0.64]	No	Yes
**Marubeni Corp	1.01	0.10	1.99	0.57	No	Yes	0.54 [0.44,0.62]	No	Yes
*Mitsubishi Chemical Holdings	1.00	0.08	1.97	0.59	No	Yes	0.46 [0.36,0.52]	Yes	Yes
**Mitsubishi Corp	1.05	0.18	1.40	0.36	Yes	No	0.29 [0.24,0.34]	Yes	No
**Mitsubishi Electric Corp	1.08	0.29	1.50	0.49	Yes	Yes	0.45 [0.39,0.52]	Yes	Yes
**Mitsubishi Estate Co Ltd	1.48	0.53	2.46	0.84	No	Yes	0.79 [0.71,0.89]	No	Yes
**Mitsubishi Heavy Industries	1.01	0.11	2.42	0.80	No	Yes	0.56 [0.45,0.66]	No	Yes
**Mitsubishi UFJ Financial Group Inc	1.00	0.03	1.44	0.32	Yes	No	0.30 [0.28,0.32]	Yes	No
**Mitsui Co Ltd	1.04	0.14	1.18	0.21	Yes	No	0.24 [0.20,0.28]	Yes	No
**Nippon Steel Sumitomo Metal Corp	1.00	0.03	1.29	0.31	Yes	No	0.30 [0.27,0.34]	Yes	No
(*)*Nissan Motor Co Ltd	1.01	0.09	2.27	0.62	No	Yes	0.50 [0.42,0.58]	Ambiguous	Yes
Nomura Holdings Inc	1.00	0.05	1.90	0.51	No	Yes	0.36 [0.33,0.40]	Yes	No
**NTT DoCoMo Inc	1.03	0.17	1.28	0.34	Yes	No	0.28 [0.24,0.32]	Yes	No
**ORIX Corp	1.06	0.23	1.23	0.33	Yes	No	0.37 [0.32,0.42]	Yes	No
**Osaka Gas Co Ltd	1.00	0.12	2.21	0.81	No	Yes	0.59 [0.47,0.70]	No	Yes
**Panasonic Corp	1.03	0.14	1.19	0.22	Yes	No	0.24 [0.20,0.28]	Yes	No
*Resona Holdings Inc	1.00	0.06	1.84	0.56	No	Yes	0.41 [0.36,0.45]	Yes	Yes
**Ricoh Co Ltd	1.05	0.25	1.23	0.29	Yes	No	0.39 [0.33,0.46]	Yes	No
**Sony Corp	1.04	0.16	1.49	0.37	Yes	No	0.26 [0.21,0.32]	Yes	No
**Sumitomo Corp	1.03	0.14	1.17	0.20	Yes	No	0.24 [0.19,0.29]	Yes	No
*Sumitomo Electric Industries	1.07	0.29	1.32	0.37	Yes	No	0.45 [0.36,0.53]	Yes	Yes
(*)*Sumitomo Mitsui Financial Group Inc	1.33	0.34	1.92	0.59	No	Yes	0.52 [0.46,0.59]	Ambiguous	Yes
**Sumitomo Mitsui Trust Holdings Inc	1.00	0.12	1.74	0.62	No	Yes	0.59 [0.48,0.67]	No	Yes
**Takeda Pharmaceutical Co	1.46	0.43	2.19	0.67	No	Yes	0.65 [0.56,0.73]	No	Yes
**Tokio Marine Holdings Inc	1.39	0.46	2.10	0.70	No	Yes	0.70 [0.61,0.76]	No	Yes
**Tokyo Gas Co Ltd	1.01	0.14	2.54	0.91	No	Yes	0.66 [0.55,0.76]	No	Yes
**Toray Industries Inc	1.02	0.14	3.24	0.98	No	Yes	0.64 [0.51,0.77]	No	Yes
Toshiba Corp	1.00	0.05	1.74	0.52	No	Yes	0.37 [0.31,0.42]	Yes	No

Table III.4: For the 48 selected stocks: Average spread and value of η in Phase 1 and Phase 2, categories in Phase 2, and predictions for η and the categories. The meaning of the stars in front of the company names is the same as for Table III.3.

CHAPTER IV

Intelligence and Randomness of Market Participants

1 Introduction

In Part I, we have seen the advantages of state-dependent approach in order book modeling, as well as its application in explaining market participants' intelligence. This new concept has shown very promising results in explaining the links between the microscopic features (such as the order arrival intensities) and the macroscopic features (such as the intraday price volatility). The Markovian framework proposed in Chapter II models the intelligence and randomness of market participants by a series of state-dependent functions (f_i,g_i in Equation II.1) and imposes very little constraints on the specific forms of these functions. This choice, while enabling the framework to include most Markovian LOB models, leaves one important practical question unanswered: how to approximate these functions in real market?

In the queue-reactive model, this problem is tackled by taking a non-parametric (point-by-point estimation) approach to approximate the order arrival intensities, in which the queue size (the state) is dicretized and truncated to a limited range. Such method requires a huge number of order book data and quickly becomes impossible when the dimension of state exceeds 3. Since many repetitive patterns have been identified in the shape of these functions on various assets, it is natural to consider the possibility of using parametric functions in describing the dynamics of order arrivals. In Chapter I, these patterns have been explained qualitatively using arguments such as the existence of priority value, the arbitrage opportunities and the adverse-selection risk, however, how to parameterize f_i and g_i so that they are capable of reproducing such patterns remains a challenging problem that we do not yet dispose a clear answer. In this chapter, we present a preliminary study of an agent-based model on the randomness and intelligence of market participants, which we believe will help understand the cause of state-dependencies in order arrival intensities and thus provide useful insights on the choice of parameterizations.

We assume the existence of three types of agents: the informed trader, the noise trader and market makers. The informed trader receives market information such as the jumps of the efficient price, which is hidden to the noise trader. He then takes advantage of this information to gain profit by sending market orders. Market makers receive also the same information but with some delay and they place limit orders as long as the average expected gain of such orders are positive (they are assumed to be risk-neutral). The informed trader and the market makers represent the "intelligent" part in order book dynamics, while the "random" part is represented by the noise trader who is assumed to send market orders according to a homogeneous Poisson process.

Interestingly, the above simple framework allows us to deduce a link between efficient price dynamics, proportion of noise trades, market volume, bid-ask spread and the equilibrium LOB state. The question of how the bid-ask spread emerges from the behavior of market participants has been discussed by many researchers. It is generally accepted that the bid-ask spread is non zero because of the existence of three types of cost: order processing costs (Huang and Stoll (1997)), inventory costs (Ho and Stoll (1981)) and adverse selection costs (Glosten and Milgrom (1985)). In our framework, both the order processing costs and the inventory costs are neglected and we consider the bid-ask spread as a purely informational phenomenon as in Glosten and Milgrom (1985). Under our settings, the bid-ask spread emerges naturally from the fact that limit orders placed too close to the efficient price have negative expected returns when being executed: the existence of the informed trader and the potential unexpected large price jumps prevent market makers from placing limit orders too close to the efficient price. Moreover, the equilibrium LOB state is shown as a direct consequence of the competition between market makers, which we believe should be closely linked with the invariant distribution of the LOB state discussed in Part I.

Although the tick value is assumed to be constant in most LOB models, it is clear that changes of the tick value will probably lead to significant changes in the microscopic features of the underlying asset. The effects of such changes are rarely discussed in the literature. In our model, the tick value is considered as one of the structural parameters. This enables the study of its roles in determining the LOB dynamics as well as the equilibrium state. The discretization of available price levels enables also a quantitative definition of the queue priority value in our framework, which is often termed only qualitatively as the advantages of an order placed on top of a queue compared to an order placed at the bottom.

This chapter is organized as follows. In Section 2, we present the core agent-based model with zero tick value. Based on the greedy assumptions of the informed trader's behavior, a link is deduced between market volume, price jump and the cumulative LOB shape. We then make an additional assumption on the zero-profit of market makers, which enables us to compute explicitly the bid-ask spread as well as the equilibrium LOB shape. In Section 3, the case of non-zero tick value is considered. We show that the constrained bid-ask spread is equal to the sum of the intrinsic bid-ask spread (without the tick value constraint) and the imposed tick value constraint. The equilibrium LOB shape under positive tick value is also deduced and we give an explicit formulation that quantifies the priority value of any order at a queue. The framework is generalized to include the market makers' uncertainty on the efficient price in Section 5. In such case, we explain how the information is propagated from the informed trader to the market makers, and the effects of this uncertainty on the bid-ask spread when a transaction happens. The speed of LOB recovery is discussed at the end of this section, in which we show that the order book will gradually recover to its equilibrium state if the reaction time of the informed trader is short enough compared to the frequency of noise trade and of information arrival.

2 Basic Model

2.1 Price Dynamics

Denote $P(t)$ the market underlying efficient price, whose dynamics is described by the following equation:

$$P(t) = P_0 + Y(t),$$

where $Y(t)$ is a compound Poisson process with jump rate λ^i and size distribution ψ (defined on $\mathbb{R}$). We will denote B as the jump size of this process: B represents the price deviation once a jump happens, which follows the distribution ψ.

We have:

$$\mathbb{E}[Y(t)] = \lambda^i t\mathbb{E}[B].$$

So for the price process $P(t)$ to be a martingale, the average value of the deviation B is assumed to be equal to zero. In such case, we have $\mathbb{V}[P(t)] = \lambda^i t\mathbb{E}[B^2]$, where the term $\lambda^i\mathbb{E}[B^2]$ represents the macroscopic volatility of the efficient price process and will be denoted by σ:

$$\sigma^2 = \lambda^i\mathbb{E}[B^2]. \tag{IV.1}$$

σ represents the fundamental volatility of the asset, due to the nature of incoming information, and can be considered as one of the invariant element in an asset's dynamics when other modifiable factors such as the tick value change.

2.2 Informed Trader, Noise Trader and Market Maker

Assume that there are three types of traders in the market:

- One informed trader: the informed trader receives the value of the price jump size B (and the efficient price $P(t)$) right before a price jump happens. He then sends his trades based on this information to gain profit. We assume that he can only send market orders.
- One noise trader: the noise trader sends random market orders to the market. We assume that these trades follows a compound Poisson process, with arrival rate λ^u and volume distribution κ^u in $\mathbb{R}$ (positive volume represents a buying order, while negative volume represents a selling order).
- Market makers: the market makers receives the value of the price jump size B (and the efficient price $P(t)$) right after a price jump happens. They place limit orders and try to make profit. We assume that they are risk neutral.

Firstly, we consider the case when the tick size is equal to 0. The LOB is composed of limit orders placed by market makers around the efficient price $P(t)$. We denote the cumulative available liquidities till $P(t)+x$ by $L(x)$. $L(x)$ represents the total quantities of the limit selling orders with price smaller or equal than $P(t)+x$ when $x>0$, and the negative of the total quantities of the limit buying orders with price larger or equal than $P(t)-x$ when $x<0$. This function $L(x)$ will be called the cumulative LOB shape function.

Unlike the modeling approach used in Madhavan, Richardson, and Roomans (1997) and Wyart, Bouchaud, Kockelkoren, Potters, and Vettorazzo (2008) where incoming order flows have non-zero impact on the concensus price's dynamics, the efficient price $P(t)$ is assumed to be independent of the order book dynamics in the above settings. In our framework, $P(t)$ should be considered as the long-term equilibrium price of the asset given current information, which changes only when new information arrives on the market. We consider the impact of order book events as the results of information propagation between market participants, which will be discussed in details in Section 5.

2.3 Some Assumptions

We first give some regularity assumptions. Let's denote the cumulative distribution function of B (the price jump size) by $F_\psi(x)$, and the cumulative distribution function of Q^u (the noise trader's trade size) by $F_{\kappa^u}(x)$.

Assumption 31. *$F_\psi(x)$ and $F_{\kappa^u}(x)$ are both differentiable. Their probability density functions will be denoted respectively by $f_\psi(x)$ and $f_{\kappa^u}(x)$. $f_\psi(x)$ and $f_{\kappa^u}(x)$ are both positive continuous functions, and for all $x \in \mathbb{R}$ (Buy-Sell Symmetry):*

$$\begin{aligned} f_\psi(-x) &= f_\psi(x) \\ f_{\kappa^u}(-x) &= f_{\kappa^u}(x). \end{aligned}$$

Moreover,

$$\begin{aligned} \mathbb{E}[|B|] &< \infty \\ \mathbb{E}[|Q^u|] &< \infty. \end{aligned}$$

For the cumulative LOB shape function $L(x)$, we assume:

Assumption 32. *The cumulative LOB shape function $L(x)$ is a continuous increasing function with $L(0) = 0$. And for all x such that $|L(x)| > 0$, $L(x)$ is differentiable, with $l(x) = L'(x)$ a continuous function in such cases.*

We can thus define the following inverse cumulative LOB shape function $L^{-1}(q)$ as:

$$L^{-1}(q) = \underset{x}{\operatorname{argmax}}\{x|L(x) = q\}. \tag{IV.2}$$

Note that $L^{-1}(q)$ is not continuous unless $L(x)$ is strictly increasing.

2.4 Links between the Trade Size Q, Price Jump B and the LOB Cumulative Shape $L(x)$

Given the function $L(x)$, we now specify the behavior of the informed trader in the following assumption. This assumption links the trade size Q^i of the informed trader with the LOB cumulative shape $L(x)$ and the price jump size B received by the informed trader.

Assumption 33. *Based on the received value of B and the cumulative LOB shape function $L(x)$ provided by the market makers, the informed trader sends his trade in a greedy way such that he wipes out all the available liquidities in the LOB till the level $P(t) + B$, thus, his trade size Q^i satisfies (Q^i is positive for a buy order and negative for a sell order), given the information signal B and the function $L(x)$:*

$$Q^i = L(B).$$

Denote $F_{\kappa^i}(x)$ the cumulative distribution function of the informed trader's trade size, $F_{\kappa^u}(x)$ the cumulative distribution function of the noise trader's trade size, and $F_\kappa(x)$ the cumulative distribution function of the trader's trade size (both of the informed and noise trader), we have the following theorem (with $r = \frac{\lambda^i}{\lambda^i+\lambda^u}$, representing the proportion of informed trader's trade, recall that $F_\psi(x)$ is the cumulative distribution function of the price jump size B, and $L^{-1}(q)$ the inverse function defined by Equation IV.2):

Theorem 1. *The cumulative distribution functions $F_{\kappa^i}(x)$, $F_{\kappa^u}(x)$ and $F_\kappa(x)$ satisfy:*

$$\begin{aligned} F_{\kappa^i}(q) &= F_{\psi^i}(L^{-1}(q)) \\ F_\kappa(q) &= rF_{\kappa^i}(q) + (1-r)F_{\kappa^u}(q). \end{aligned} \tag{IV.3}$$

2.5 The Bid-Ask Spread and the Equilibrium LOB Shape

Now the question remains for the market makers to determine the function $L(x)$, given the greedy behavior of the informed trader and the random behavior of the noise trader.

Let's consider the profit of passive selling orders placed between $P(t)+x$ and $P(t)+x+\delta p$ for some $x, \delta p > 0$, given the fact that these orders are totally executed. This conditional expected gain will be denoted by $G(x,\delta p)$, and we have the following equation (with v a random variable that is equal to 1 if the trade is initiated by the informed trader, 0 if it is initiated by the noise trader):

$$\begin{aligned} G(x,\delta p) &= \int_x^{x+\delta p} l(s)s\,ds - [L(x+\delta p) - L(x)]\mathbb{E}[vB|Q \geq L(x+\delta p)] \\ &= \int_x^{x+\delta p} l(s)s\,ds - [L(x+\delta p) - L(x)]r\frac{1-F_{\kappa^i}(x+\delta p)}{1-F_\kappa(L(x+\delta p))}\mathbb{E}[B|B > x+\delta p]. \end{aligned}$$

If $L(x) > 0$ and $L(x+\delta p) > L(x)$ for δp sufficiently small (or equivalently that for all $\delta p > 0$, $\sup_{s\in[x,x+\delta p]} l(s) > 0$, we can define the average profit per unity at x, denoted by $G(x)$:

$$\begin{aligned} G(x) &= \lim_{\delta p\to 0^+} \frac{G(x,\delta p)}{L(x+\delta p)-L(x)} \\ &= x - r\frac{1-F_{\kappa^i}(L(x))}{1-F_\kappa(L(x))}\mathbb{E}[B|B > x] \\ &= x - \frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{1-F_\kappa(L(x))} \\ &= x - \frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{1-rF_\psi(x)-(1-r)F_{\kappa^u}(L(x))}. \end{aligned}$$

It is easy to see that $G(x)$ is a decreasing function of $L(x)$, so that when $G(x)$ is positive, the market makers will add new liquidities to the LOB (increase the value of $L(x)$), as they can still make profit in average. This kind of competition between market makers results in the following zero profit assumption:

Assumption 34. *For every $x \in \mathbb{R}^+$, if $L(x) > 0$ and $l(x) > 0$, then the conditional expected average gain per unity of passive orders placed at x, given the fact that they are totally executed, is equal to 0, that is, when $L(x) > 0$ and $l(x) > 0$:*

$$G(x) = 0.$$

Equivalently, we have:

$$x = r\frac{\mathbb{E}[B\mathbf{1}_{B>x}]}{1-F_\kappa(L(x))}. \tag{IV.4}$$

The case corresponding to passive buying orders can be similarly obtained:

Assumption 35. *For every $x \in \mathbb{R}^-$, if $L(x) < 0$ and $l(x) > 0$, then the conditional expected average gain per unity of passive orders placed at x, given the fact that they are totally executed, is equal to* 0*, that is, when $L(x) < 0$ and $l(x) > 0$:*

$$G(x) \quad = \quad 0.$$

Equivalently, we have:

$$x = r\frac{\mathbb{E}[B\mathbf{1}_{B<x}]}{F_\kappa(L(x))}. \tag{IV.5}$$

Remark 1. *In the continuous case, it might be difficult to image how the competitions between different market makers happen. One can think of the case that every market maker specifies his own $L(x)$ (cumulative liquidities that he provides till each price level), then Assumption 34 means that, when there is still space for future profit in x ($G(x) > 0$), other market makers will come to the market, and increase the liquidities in the LOB so that in the end, no further profit can be made by adding liquidities anywhere.*

Remark 2. *The above zero profit assumptions can also be seen as the generalized version of the zero profit assumption proposed in Glosten and Milgrom (1985), in which the zero profit is only assumed for the two best offer limits. It is also interesting to point out that, under this generalized zero profit assumption, those fast market makers can still make profit, as their orders are placed earlier in the LOB (this point will be made clearer in the discrete case).*

Now considering the ask side of the limit order book. If $L(x) > 0$ for all $x > 0$, we shall have:

$$\begin{aligned}
\lim_{x\to 0^+} G(x) &= -r\frac{\mathbb{E}[B\mathbf{1}_{B>0}]}{1-F_\kappa(0)} \\
&= -2r\mathbb{E}[B\mathbf{1}_{B>0}] \\
&< 0.
\end{aligned}$$

As $G(x)$ is a continuous function of x, there exists some $\delta x > 0$, such that for all $x \in (0, \delta x)$, $G(x) < 0$. Then according to Assumption 34, we should have $l(x) = 0$ for all $x \in (0, \delta x)$, which means that $L(x) = 0$ for all $x \in (0, \delta x)$. This is clearly controversial.

Now assume that $L(x) = 0$ till some positive value η, and that $L(x) > 0$, for all $x > \eta$. From Assumption 34, we have, for all $x > \eta$, if $l(x) > 0$:

$$x \quad = \quad \frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{1-rF_{\psi^i}(x)-(1-r)F_{\kappa^u}(L(x))}.$$

We shall have thus, for all $x > \eta$, if $l(x) > 0$:

$$
\begin{aligned}
L(x) &= F_{\kappa^u}^{-1}\left(\frac{1 - rF_{\psi^i}(x)}{1-r} - \frac{r\mathbb{E}[B\mathbf{1}_{B>x}]}{(1-r)x}\right) \\
&= F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\right).
\end{aligned}
\tag{IV.6}
$$

Now if there exists some $x' > \eta$ and $\delta x' > 0$, such that $l(x) = 0$ for all $x \in [x', x' + \delta x']$. Let's define $x_r = \inf\{x | l(x) > 0, x > x' + \delta x'\}$ and $x_l = \sup\{x | l(x) > 0, x < x'\}$. If $x_r < \infty$, then by the continuity of $l(x)$ and $L(x)$, we have:

$$
\begin{aligned}
L(x_l) &= F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x_l}, 1)]\right) \\
L(x_r) &= F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x_r}, 1)]\right).
\end{aligned}
$$

By Assumption 31, we have $L(x_r) > L(x_l)$. While by the fact that $l(x) = 0$ for all $x \in [x_l, x_r]$, we should have $L(x_r) = L(x_l)$. These two results are clearly controversial. So either we have $l(x) = 0$ for all $x \geq x_l$, or such x' and $\delta x'$ do not exist. For the first case, we should have $L(x) = F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)])$ for all $x \in (\eta, x_l)$, thus $l(x_l) > 0$ by the continuity of $f_{\kappa^u}(x)$, which is controversial to the fact that $l(x_l) = 0$. Thus for any $x' > \eta$, and any $\delta x' > 0$, one can always find a $x \in [x', x' + \delta x']$ such that $l(x) > 0$. Then by the continuity of the function $L(x)$, we have that Equation IV.6 holds for all $x > \eta$.

$L(x)$ is thus a strictly increasing function when $x \geq \eta$, and when $x \to \eta^-$:

$$
\lim_{x \to \eta^-} L(x) = F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{\eta}, 1)]\right) = 0.
$$

Which means that:

$$
\begin{aligned}
1 - r\mathbb{E}[\max(\frac{B}{\eta}, 1)] &= 0.5(1-r) \\
\frac{1+r}{2r} &= \mathbb{E}[\max(\frac{B}{\eta}, 1)].
\end{aligned}
$$

The bid side case can be similarly obtained. The above arguments give the following theorem:

Theorem 2. *By Assumption 31, 32, 33, 34 and 35, the cumulative LOB shape function is uniquely determined and satisfies $L(x) = -L(-x)$. Moreover, we have, $L(x) = 0$, for $x \in [-\eta, \eta]$, where η is the unique solution of the following equation:*

$$
\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\eta}, 1)]. \tag{IV.7}
$$

And for $x > \eta$:

$$
L(x) = F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{x}, 1)]\right),
$$

for $x < -\eta$:

$$
L(x) = -F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{-x}, 1)]\right).
$$

In particular, the intrinsic bid-ask spread satisfies $\phi = 2\eta$.

2.6 Variance per Trade

Consider now the price variance per trade (σ_{tr}) in our model. We have (Let's P_i denotes the value of the efficient price after the i-th event [1], and τ the number of events till the first trade after the i-th event):

$$\begin{aligned}
\sigma_{tr}^2 &= \mathbb{E}[(P_{i+\tau} - P_i)^2] \\
&= \sum_{j=1}^{\infty} \mathbb{P}[\tau = j]\mathbb{E}[(P_{i+j} - P_i)^2 | \tau = j] \\
&= \mathbb{E}[B^2 | B \leq \eta] \sum_{j=1}^{\infty} \mathbb{P}[\tau = j](j-1) \\
&\quad + \frac{r\mathbb{P}[|B| > \eta]\mathbb{E}[B^2 | B > \eta]}{1 - r\mathbb{P}[|B| \leq \eta]} \\
&= \frac{r\mathbb{E}[B^2 \mathbf{1}_{|B|\leq\eta}] + r\mathbb{E}[B^2 \mathbf{1}_{|B|>\eta}]}{1 - r\mathbb{P}[|B| \leq \eta]} \\
&= \frac{r\mathbb{E}[B^2]}{1 - r\mathbb{P}[|B| \leq \eta]}.
\end{aligned}$$

By Equation IV.7, we have:

$$1 - r\mathbb{P}[|B| \leq \eta] = \frac{r\mathbb{E}[|B|\mathbf{1}_{|B|>\eta}]}{\eta}.$$

Thus,

$$\sigma_{tr}^2 = \frac{\mathbb{E}[B^2]\eta}{\mathbb{E}[|B|\mathbf{1}_{|B|>\eta}]}.$$

For $\sigma_{tr}^2 \sim \eta^2$ (that is to have the linear relationship between the bid-ask spread and the volatility per trade), one must have: $\mathbb{E}[|B|\mathbf{1}_{|B|>\eta}] \sim \eta^{-1}$, which implies actually a power-low distribution of order -3 on the price jump size B.

3 Tick Size

In this section, we study the effects of introducing the tick size, denoted by α, that constraints the minimal price change unit. The same price dynamics described in the previous section still applies, but the cumulative LOB shape becomes now a piece-wise constant function $L(x)$. Due to the price discretization, the discontinuity points of $L(x)$ actually depends on the relative position of the current efficient price $P(t)$ in the price grid. To deal with this, the following notations will be used in this section. Let's denote the smallest possible price level that is greater than or equal to the current efficient price $P(t)$ by $\bar{P}(t)$, and their relative distance by $d := \bar{P}(t) - P(t)$. We have $d \in [0, \alpha)$, and the cumulative LOB shape function $L(x)$ in such case will be denoted by $L^d(x)$ from now on, with $l^d(i)$ representing the quantities placed at the i−th limit ($i = 1$ represents the closest price level that is larger or equal to $P(t)$, while $i = -1$ represents the closest price level that is smaller than $P(t)$).

[1] An event can be either a trade sent by noise trader, or an information update B which may or may not trigger a trade, depending on whether $|B| > |\eta|$ or not.

Assumption 31 still applies in this case as well as the greedy behavior of the informed trader (Assumption 33). For the cumulative LOB shape function $L^d(x)$, we assume now:

Assumption 36. *The cumulative LOB shape function $L^d(x)$ is an increasing piece-wise constant function, with discontinuities at $x = d + i\alpha$ for $i \in \mathbb{Z}$.*

Given the value of d and the LOB shape $L^d(x)$, the trade size Q^i of the informed trader when a new price deviation information B is received becomes a discrete value random variable, for which we can write the probability function as (for $d > 0$):

$$\begin{aligned} P^d_{\kappa^i}(Q^i = L^d(d+(i-1)\alpha)) &= P(B \in (d+(i-1)\alpha, d+i\alpha]), \text{ for } i > 0 \\ P^d_{\kappa^i}(Q^i = L^d(d+i\alpha)) &= P(B \in [d+(i-1)\alpha, d+i\alpha)), \text{ for } i < 0. \end{aligned}$$

3.1 Constrained Bid-Ask Spread

Let's compute again the average gain of passive selling orders placed at different limits. We have, when $l(i)$ (the number of limit orders placed at the i-th limit) > 0, $G^d(i)$ (the expected average gain of passive orders placed at the price level $d + (i-1)\alpha$ given the fact that they are completely executed, for $i \in \mathbb{N}^+$) writes as:

$$\begin{aligned} G^d(i) &= d+(i-1)\alpha - \mathbb{E}[\nu B | Q \geq \sum_{j=1}^{i} l(j)] \\ &= d+(i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]}{1-F_\kappa(L^d(d+(i-1)\alpha))}. \end{aligned}$$

$G^d(i)$ can be understood as the expected gain of a newly inserted infinitesimal limit order at the i−th limit, under the condition that it is executed by some market order. It is natural to use the following zero profit assumption of passive orders at each price level:

Assumption 37. *For every $i \in \mathbb{N}^+$, if $l^d(i) > 0$, then the conditional expected average gain of passive selling orders placed at $d+(i-1)\alpha$, given the fact that they are totally executed, is equal to 0, that is, when $l^d(i) > 0$:*

$$G^d(i) = 0.$$

Equivalently, we have:

$$d+(i-1)\alpha = \frac{r\mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]}{1-F_\kappa(L^d(d+(i-1)\alpha))}. \tag{IV.8}$$

When $l^d(i) = 0$, we can define $\tilde{G}^d(i)$, the potential conditional expected average gain per unity of passive orders placed at $d+(i-1)\alpha$ by the following equation:

$$\begin{aligned} \tilde{G}^d(i) &= d+(i-1)\alpha - \lim_{\delta v \to 0+} \frac{r\mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]}{1-F_\kappa(\max(0, L^d(d+(i-2)\alpha)+\delta v))} \\ &= d+(i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B\geq d+(i-1)\alpha}]}{1-rF_\psi(d+(i-1)\alpha)-(1-r)F_{\kappa^u}(\max(0, L^d(d+(i-2)\alpha)))}. \end{aligned}$$

One can image that when the potential gain $\tilde{G}^d(i)$ is positive, market makers will place passive orders at the i−th limit. This idea gives the following assumption:

Assumption 38. *For every $i \in \mathbb{N}^+$, if $l^d(i) = 0$, then the potential conditional expected average gain of passive selling orders placed at $d + (i-1)\alpha$, given the fact that they are totally executed, is less than or equal to* 0*, that is, when $l^d(i) = 0$:*

$$\tilde{G}^d(i) \leq 0.$$

Equivalently, we have:

$$d + (i-1)\alpha \leq \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(i-1)\alpha}]}{1 - rF_\psi(d + (i-1)\alpha) - (1-r)F_{\kappa^u}(\max(0, L^d(d + (i-2)\alpha)))}.$$

Remark 1. *We will see latter that $\tilde{G}^d(i)$ represents also the priority value of the first order placed at the i-th limit.*

As for the bid side, we have the following two corresponding assumptions:

Assumption 39. *For every $i \in \mathbb{N}^-$, if $l^d(i) < 0$, then the conditional expected average gain of passive buying orders placed at $d + i\alpha$, given the fact that they are totally executed, is equal to* 0*, that is, when $l^d(i) < 0$:*

$$G^d(i) = 0.$$

Equivalently, we have:

$$d + i\alpha = \frac{r\mathbb{E}[B\mathbf{1}_{B \leq d+i\alpha}]}{F_\kappa(L^d(d + i\alpha))}. \tag{IV.9}$$

Assumption 40. *For every $i \in \mathbb{N}^-$, if $l^d(i) = 0$, then the potential conditional expected average gain of passive buying orders placed at $d + i\alpha$, given the fact that they are totally executed, is less than or equal to* 0*, that is, when $l^d(i) = 0$:*

$$\tilde{G}^d(i) \leq 0.$$

Equivalently, we have:

$$d + i\alpha \geq \frac{r\mathbb{E}[B\mathbf{1}_{B \leq d+i\alpha}]}{rF_\psi(d + i\alpha) + (1-r)F_{\kappa^u}(\min(0, L^d(d + (i+1)\alpha)))}.$$

As in the continuous case, let's consider the first non empty ask limit k_r^d, we should have:

$$\begin{aligned} d + (k_r^d - 1)\alpha &= \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(k_r^d-1)\alpha}]}{1 - F_\kappa(l^d(k_r^d))} \\ d + (k_r^d - 1)\alpha &> \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(k_r^d-1)\alpha}]}{1 - F_\kappa(0)}, \end{aligned}$$

which leads to:

$$F_{\kappa^u}(l^d(k_r^d)) = \frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d + (k_r^d - 1)\alpha}, 1)].$$

Thus for $l^d(k_r^d)$ to be positive, we should have $\frac{1}{1-r} > 0.5 + \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+(k_r^d-1)\alpha}, 1)]$, which means:

$$\frac{1+r}{2r} > \mathbb{E}[\max(\frac{B}{d+(k_r^d-1)\alpha}, 1)]. \tag{IV.10}$$

Equation IV.10 gives the necessary and sufficient condition for k_r^d. Compared with Equation IV.7, we have:

$$k_r^d = \min\{k \in \mathbb{N}^+ | d + (k-1)\alpha > \eta\},$$

with the operator $\lceil x \rceil$ defined as the smallest integer that is strictly larger than x, we have furthermore:

$$k_r^d = 1 + \lceil \frac{\eta - d}{\alpha} \rceil.$$

Similarly, for the first non-empty limit at the bid side, we shall have:

$$k_l^d = \lceil \frac{\eta + d}{\alpha} \rceil.$$

Thus the conditional constrained bid-ask spread ϕ_α^d, given the value of d, satisfies:

$$\phi_\alpha^d = \alpha(\lceil \frac{\eta - d}{\alpha} \rceil + \lceil \frac{\eta + d}{\alpha} \rceil).$$

Assuming d is uniformly distributed between $[0, \alpha)$, we can compute the average value of the constrained bid-ask spread by integrating ϕ_α^d:

$$\phi_\alpha = \int_0^{\alpha^-} \lceil \frac{\eta - s}{\alpha} \rceil + \lceil \frac{\eta + s}{\alpha} \rceil ds.$$

Denote $u := \frac{\eta}{\alpha}$, we have:

$$\phi_u = \alpha \int_0^{1^-} \lceil u - x \rceil + \lceil u + x \rceil dx.$$

Write $u = u_i + u_f$, where u_i represents the integral part of u, we have furthermore:

$$\begin{aligned}
\phi_\alpha &= \alpha \int_0^{1^-} \lceil u_i + u_f - x \rceil + \lceil u_i + u_f + x \rceil dx \\
&= \alpha[\int_0^{u_f} (u_i + 1)dx + \int_{u_f^+}^{1^-} u_i dx + \int_0^{1-u_f} (u_i + 1)dx + \int_{(1-u_f)^+}^{1^-} (u_i + 2)dx] \\
&= \alpha[u_f(u_i + 1) + (1 - u_f)u_i + (1 - u_f)(u_i + 1) + u_f(u_i + 2) \\
&= \alpha[2u_i + 2u_f + 1] \\
&= \alpha + 2\eta \\
&= \alpha + \phi.
\end{aligned}$$

The above arguments are summarized in the following theorem, which proves that the discrete spread is a linearly increasing function of the tick size α in our framework:

Theorem 3. *The average constrained spread ϕ_α when the tick size is equal to α, satisfies the following equation:*

$$\phi_\alpha = \alpha + \phi, \tag{IV.11}$$

where ϕ is the intrinsic bid-ask spread of the asset when the tick value is equal to 0.

Also, the following theorem on the cumulative LOB shape can be established:

Theorem 4. *By Assumption 33, 36, 37, 38, 39 and 40, the cumulative LOB shape function is uniquely determined. We have, $l^d(i) = 0$ for all $-k_l^d < i < k_r^d$, where k_l^d and k_r^d are two positive integers determined by the following equations:*

$$\begin{aligned} k_r^d &= 1 + \lceil \frac{\eta - d}{\alpha} \rceil \\ k_l^d &= \lceil \frac{\eta + d}{\alpha} \rceil, \end{aligned}$$

with η the unique solution of the following equation:

$$\frac{1+r}{2r} = \mathbb{E}[\max(\frac{B}{\eta}, 1)].$$

And for $h \geq k_r^d$:

$$L^d(d + (h-1)\alpha) = F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{d+(h-1)\alpha}, 1)]),$$

for $h \leq -k_l^d$:

$$L^d(d + h\alpha) = -F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{-d-h\alpha}, 1)]).$$

3.2 Daily Volume

Now let's consider the daily volume V before and after the introduction of the tick size (denote the latter by V_α). We have, the expected value of V satisfies:

$$\mathbb{E}[V] = \lambda^i \mathbb{E}[|Q^i|] + \lambda^u \mathbb{E}[|Q^u|].$$

In the previous discussion, we implicitly assume that the distribution of the noise trader's bet size does not depend on the cumulative LOB shape function $L(x)$, while the informed trader's bet size distribution is linked with $L(x)$ by Assumption 33. Note that from Theorem 2, we have an explicit expression of $L(x)$. We have thus:

$$\begin{aligned} \mathbb{E}[|Q^i|] &= 2\mathbb{P}(Q^i > 0)\mathbb{E}[Q^i | Q^i > 0] \\ &= 2\mathbb{E}[L(B^i)\mathbf{1}_{B_i > a}] \\ &= 2\int_a^\infty F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{b}, 1)]) f_\psi(b) db. \end{aligned}$$

When the tick size is introduced, we shall have:

$$\mathbb{E}[V_\alpha] = \lambda^i \mathbb{E}[|Q_\alpha^i|] + \lambda^u \mathbb{E}[|Q^u|].$$

Where $\mathbb{E}[|Q_\alpha^i|]$ satisfies:

$$\begin{aligned} \mathbb{E}[|Q_\alpha^i|] &= 2\mathbb{E}[Q^i \mathbf{1}_{Q^i > 0}] \\ &= \frac{2}{\alpha}\int_0^{\alpha^-} \mathbb{E}[Q_d^i \mathbf{1}_{Q_d^i > 0}] dd. \end{aligned}$$

We have, for $d \neq 0$:

$$\mathbb{E}[Q_d^i \mathbf{1}_{Q_d^i > 0}] \quad = \quad \sum_{h=k_r^d}^{\infty} L^d(h)(F_\psi(d + h\alpha) - F_\psi(d + (h-1)\alpha).$$

Thus $\mathbb{E}[V_\alpha]$ is a decreasing function of α, and we have:

$$\lim_{\alpha \to \infty} \mathbb{E}[V_\alpha] = \lambda^u \mathbb{E}[|Q^u|].$$

3.3 Priority Value

Assumption 37 and 39 tell that the profit of the limit order placed at the bottom of each queue is equal to 0. Market makers may still make profit if their orders are placed before. The expected profit of the limit order placed on top of each limit is actually equal to the potential average gain $\tilde{G}^d(i)$ when $L(x)$ is given. Considering the profit of a limit selling order placed on top of the $i-$th limit, we have, for $i \geq k_r^d$:

$$\begin{aligned}
\tilde{G}^d(i) &= d + (i-1)\alpha - \lim_{\delta v \to 0+} \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(i-1)\alpha}]}{1 - F_\kappa(\max(0, L^d(d + (i-2)\alpha) + \delta v))} \\
&= d + (i-1)\alpha - \frac{r\mathbb{E}[B\mathbf{1}_{B \geq d+(i-1)\alpha}]}{1 - rF_\psi(d + (i-1)\alpha) - (1-r)F_{\kappa^u}(\max(0, L^d(d + (i-2)\alpha)))}.
\end{aligned}$$

This gives us the following theorem (formulas on the priority value of limit buying orders are similar and omitted here):

Theorem 5. *The priority value at the $i-$th limit can be written as:*
For $i = k_r^d$,

$$\begin{aligned}
\tilde{G}^d(i) &= \mathbb{E}[B\mathbf{1}_{B \geq d+(k_r^d-1)\alpha}]\{\frac{1}{\mathbb{E}[\max(\frac{B}{d+(k_r^d-1)\alpha}, 1)] - F_\psi(d + (k_r^d - 1)\alpha)} \\
&\quad - \frac{1}{\frac{1+r}{2r} - F_\psi(d + (k_r^d - 1)\alpha)}\},
\end{aligned} \tag{IV.12}$$

for $i > k_r^d$,

$$\begin{aligned}
G_p^d(i) &= \mathbb{E}[B\mathbf{1}_{B \geq d+(i-1)\alpha}]\{\frac{1}{\mathbb{E}[\max(\frac{B}{d+(i-1)\alpha}, 1)] - F_\psi(d + (i-1)\alpha)} \\
&\quad - \frac{1}{\mathbb{E}[\max(\frac{B}{d+(i-2)\alpha}, 1)] - F_\psi(d + (i-1)\alpha)}\}.
\end{aligned}$$

4 Examples

4.1 Power-law Distributed Information

Assume that the absolute value of the price jump B follows a power-law distribution, that is: $f_\psi(x) = Cx^{-\gamma}$ when $|x| > \epsilon > 0$. Note that we should have $\gamma > 3$ to have a finite variance in price. The tick value is set to zero in the this example.

Fundamental Bid-Ask Spread

We have, in this case:

$$\begin{aligned}
\frac{1+r}{2r} &= \mathbb{E}[\max(\frac{B}{\eta},1)] \\
&= \frac{\mathbb{E}[\max(B,\eta)]}{\eta} \\
&= \frac{\eta(1-\frac{C}{\gamma-1}\eta^{-\gamma+1})+\frac{C}{\gamma-2}\eta^{-\gamma+2}}{\eta} \\
&= (1-\frac{C}{\gamma-1}\eta^{-\gamma+1})+\frac{C}{\gamma-2}\eta^{-\gamma+1} \\
\frac{1-r}{2rC} &= \frac{1}{(\gamma-1)(\gamma-2)\eta^{\gamma-1}} \\
\eta &= (\frac{2rC}{(1-r)(\gamma-1)(\gamma-2)})^{\frac{1}{\gamma-1}}.
\end{aligned}$$

η and the ratio between the intensity of price jump and noise trade ($\frac{\lambda^i}{\lambda^u}$) are related by the following relationship:

$$\eta \sim (\frac{\lambda^i}{\lambda^u})^{\frac{1}{\gamma-1}}.$$

Cumulative LOB Shape

For the cumulative LOB shape function $L(x)$, we have:

$$\mathbb{E}[\max(\frac{B}{x},1)] = \frac{C}{(\gamma-1)(\gamma-2)x^{\gamma-1}},$$

which means:

$$L(x) = F_{\kappa^u}^{-1}(\frac{1}{1-r} - \frac{rC}{(1-r)(\gamma-1)(\gamma-2)x^{\gamma-1}}).$$

We can see that if $L(x) \sim x^{\beta}$, we shall have: $f_{\kappa^u} \sim x^{\frac{-\gamma+1}{\beta}}$ and vice versa.

5 Generalization

Till now, market makers are assumed to obtain the exact information on the efficient price once after it has been known to the informed trader. This is obviously unrealistic as it neglects one important risk faced by market makers when placing limit orders: the uncertainty on the efficient price. In this section, we will study the case when such information is no longer available for market makers.

5.1 How Information is Digested

Let's reconsider the case of zero tick value. At the moment t, a buying transaction happens, let's assume that market makers hold only the exact information on the efficient price right before the transaction (denoted by $P(t-)$, we set $P(t-)=0$ without loss of generality), but no longer know the value of the efficient price after this transaction, that is, market makers no longer obtain the information of the price jump B. Given a buying market order (we consider only the ask side of the LOB in the sequel for simplicity) of volume q (till the price level p') at time t, assume now, that if this trade is issued by the informed trader, then after this trade, he is going to take any liquidities below the new efficient price p' with some reaction time, distributed exponentially with rate r^a. In such case, we have, immediately after a transaction, the gain of limit orders placed between x and $x+\delta p$ can be written as (v^i is a random variable with $v^i=1$ if the transaction comes from the informed trader, $v^i=0$ if the transaction comes from the noise trader):

$$g^{p'}(x,\delta p) \quad = \quad v^i \tilde{g}^{p'}(x,\delta p) + (1-v^i) g(x,\delta p),$$

where $\tilde{g}^{p'}(x,\delta p)$ represents the gain in the case of an informed trade, and $g(x,\delta p)$ that in the case of a noise trade.

Three different cases can happen when the trade is initiated by the informed trader given the fact that the limit orders placed between between x and $x+\delta p$ are completely executed (and no other event happens in-between):

- e^1: The limit orders are consumed by a noise trade.
- e^2: The limit orders are consumed by an informed trade, due to the arrival of a new price jump.
- e^3: The limit orders are consumed by an informed trade, without the arrival of a new price jump.

For $\tilde{g}^{p'}(x,\delta p)$, we have (define $\mathbf{1}_{e^i}=1$ if event e^i happens and $=0$ otherwise, and denote by B^{new} the new price jump):

$$\begin{aligned}
\tilde{g}^{p'}(x,\delta p) &= \mathbf{1}_{e^1}\int_x^{x+\delta p} l(s)(s-p')ds + \mathbf{1}_{e^2}[\int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))(B^{new}+p')] \\
&\quad + \mathbf{1}_{e^3}[\int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))p'] \\
&= \int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))p' - \mathbf{1}_{e^2}(L(x+\delta p)-L(x))B^{new}.
\end{aligned}$$

As for $g(x,\delta p)$, the event e^3 is not possible, we have:

$$\begin{aligned}
g(x,\delta p) &= \mathbf{1}_{e^1}\int_x^{x+\delta p} l(s)(s)ds + \mathbf{1}_{e^2}[\int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))B^{new}] \\
&= \int_x^{x+\delta p} l(s)sds - \mathbf{1}_{e^2}(L(x+\delta p)-L(x))B^{new}.
\end{aligned}$$

We then compute the conditional expectations of these two gains.

$$\begin{aligned}
\tilde{G}^{p'}(x,\delta p) &= \mathbb{E}[\tilde{g}^{p'}(x,\delta p)|v^i=1] \\
&= \int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))p' \\
&\quad -(L(x+\delta p)-L(x))\mathbb{E}[\mathbf{1}_{e^2}B^{new}|v^i=1].
\end{aligned}$$

For $x < p'$ and $x + \delta p < p'$, we have (denote $r^i = \frac{\lambda^i}{\lambda^i+\lambda^u+\lambda^a}$ and $r^u = \frac{\lambda^u}{\lambda^i+\lambda^u+\lambda^a}$):

$$\begin{aligned} G^{p'}(x,\delta p) &= \int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))p' \\ &\quad -(L(x+\delta p)-L(x))\frac{r^i\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x+\delta p-p'}]}{1-r^iF_{\kappa^i}(L(x+\delta p))-r^uF_{\kappa^u}(L(x+\delta p))}. \end{aligned}$$

For $x \geq p'$, we have:

$$\begin{aligned} \tilde{G}^{p'}(x,\delta p) &= \int_x^{x+\delta p} l(s)sds - (L(x+\delta p)-L(x))p' \\ &\quad -(L(x+\delta p)-L(x))\frac{r\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x+\delta p-p'}]}{1-rF_{\kappa^i}(L(x+\delta p))-(1-r)F_{\kappa^u}(L(x+\delta p))}. \end{aligned}$$

And:

$$\begin{aligned} G(x,\delta p) &= \mathbb{E}[g(x,\delta p)|v^i=0] \\ &= \int_x^{x+\delta p} l(s)sds - [L(x+\delta p)-L(x)]\frac{r\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x+\delta p}]}{1-rF_\psi(x+\delta p)-(1-r)F_{\kappa^u}(L(x+\delta p))}. \end{aligned}$$

Thus, we have, if $L(x) > 0$ and $L(x+\delta p) > L(x)$ for all δp sufficiently small, the average expected gain at the price x writes:

For $x < p'$:

$$\begin{aligned} \tilde{G}^{p'}(x) &= \lim_{\delta p\to 0^+}\frac{\tilde{G}^{p'}(x,\delta p)}{L(x+\delta p)-L(x)} \\ &= x-p'-\frac{r^i\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x-p'}]}{1-r^iF_\psi(x-p')-r^uF_{\kappa^u}(L(x))}. \end{aligned}$$

For $x \geq p'$, we have:

$$\begin{aligned} \tilde{G}^{p'}(x) &= \lim_{\delta p\to 0^+}\frac{\tilde{G}^{p'}(x,\delta p)}{L(x+\delta p)-L(x)} \\ &= x-p'-\frac{r\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x-p'}]}{1-rF_\psi(x-p')-(1-r)F_{\kappa^u}(L(x))}. \end{aligned}$$

And:

$$\begin{aligned} G(x) &= \lim_{\delta p\to 0^+}\frac{G(x,\delta p)}{L(x+\delta p)-L(x)} \\ &= x-\frac{r\mathbb{E}[B^{new}\mathbf{1}_{B^{new}>x}]}{1-rF_\psi(x)-(1-r)F_{\kappa^u}(L(x))}. \end{aligned}$$

Finally, we have, (denote $L^-(p')$ as the cumulative LOB quantities right before the transaction at the price level p' and denote $r^{p'} = \frac{r(1-F_\psi(p'))}{1-rF_\psi(p')-(1-r)F_{\kappa^u}(L^-(p'))}$:

$$G^{p'}(x) \quad = \quad r^{p'}\tilde{G}^{p'}(x) + (1 - r^{p'})G(x). \tag{IV.13}$$

Note that Theorem 2 still applies for $L^-(p')$, as the exact value of the efficient price right before the transaction is assumed to be known to the market makers. $r^{p'}$ can thus be simplified as:

$$r^{p'} \quad = \quad \frac{p'}{\mathbb{E}[B^{new}|B^{new} > p']}.$$

Using the above equation, we have, for $x < p'$, since B^{new} follows the save law as B :

$$G^{p'}(x) \quad = \quad x - r^{p'}p' - \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - rF_\psi(x) - (1-r)F_{\kappa^u}(L(x))} - \frac{r^{p'}r^i\mathbb{E}[B\mathbf{1}_{B>x-p'}]}{1 - r^iF_\psi(x-p') - r^uF_{\kappa^u}(L(x))},$$

and for $x \geq p'$:

$$G^{p'}(x) \quad = \quad x - r^{p'}p' - \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - rF_\psi(x) - (1-r)F_{\kappa^u}(L(x))} - \frac{r^{p'}r\mathbb{E}[B\mathbf{1}_{B>x-p'}]}{1 - rF_\psi(x-p') - (1-r)F_{\kappa^u}(L(x))}.$$

Similar to Assumption 34, the competition between market makers results in the following generalized version of the zero-profit assumption:

Assumption 41. *For any $x > 0$, if $L(x) > 0$ and $l(x) > 0$, then the average conditional expected gain $G^c(x)$, defined as*

$$G^c(x) = \lim_{\delta p \to 0^+} \frac{G^c(x, \delta p)}{L(x+\delta p) - L(x)},$$

satisfies:

$$G^c(x) = 0,$$

where $G^c(x, \delta p)$ is the conditional expected gain of passive orders placed between $P(t) + x$ and $P(t) + x + \delta p$, given the facts that these orders are completely consumed by a market order.

Assumption 41 all together with the above arguments give the following theorem concerning the equilibrium cumulative LOB shape immediately after a transaction:

Theorem 6. *If the informed trader is assumed to react to the new profitable liquidities with some exponentially distributed reaction time (with intensity λ^a) and the efficient price $P(t-)$ is assumed to be known to the market makers, then under Assumption 31, 33 and 41, immediately after a buying transaction happens at the price $P(t-) + p'$ at time t, the cumulative LOB shape function satisfies: for every $x \in \mathbb{R}^+$, if $L(x) > 0$ and $l(x) > 0$, then $G^{p'}(x) = 0$, that is:*

for $x < p'$:

$$x - r^{p'}p' \quad = \quad \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - rF_\psi(x) - (1-r)F_{\kappa^u}(L(x))} + \frac{r^{p'}r^i\mathbb{E}[B\mathbf{1}_{B>x-p'}]}{1 - r^iF_\psi(x-p') - r^uF_{\kappa^u}(L(x))},$$

for $x \geq p'$:

$$x - r^{p'}p' \quad = \quad \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - rF_\psi(x) - (1-r)F_{\kappa^u}(L(x))} + \frac{r^{p'}r\mathbb{E}[B\mathbf{1}_{B>x-p'}]}{1 - rF_\psi(x-p') - (1-r)F_{\kappa^u}(L(x))}.$$

Consider now the new best ask price $a^{p'}$ immediately after the transaction at price p'. We shall have, if $a^{p'} < p'$:

$$\begin{aligned}(a^{p'} - r^{p'} p') &= \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>a^{p'}}]}{1 - rF_\psi(a^{p'}) - (1-r)/2} \\ &\quad + \frac{r^{p'} r^i \mathbb{E}[B\mathbf{1}_{B>a^{p'}-p'}]}{1 - r^i F_\psi(a^{p'} - p') - r^u/2}.\end{aligned}$$

To simplify our analysis, we make the following assumption on the informed trader's reaction speed:

Assumption 42. *The reaction rate λ^a of the informed trader is much larger than the information arrival rate λ^i and the noise trade arrival rate λ^u, that is:*

$$\begin{aligned} r^i &\approx 0 \\ r^u &\approx 0. \end{aligned}$$

We have, under the above assumption:

$$(a^{p'} - r^{p'} p') \approx \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>a^{p'}}]}{1 - rF_\psi(a^{p'}) - (1-r)/2}.$$

We can see that when $a^{p'} = p'$, $(a^{p'} - r^{p'} p') > \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>a^{p'}}]}{1-rF_\psi(a^{p'})-(1-r)/2}$ by the fact that $p' > \eta$, so indeed $a^{p'} < p'$ when λ^a is sufficiently large compared to λ^u and λ^i.

We have the following two approximations:

$$(a^{p'} - r^{p'} p')\frac{1+r}{2r} \approx (1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>a^{p'}}] + (a^{p'} - r^{p'} p')F_\psi(a^{p'}),$$

$$\frac{1+r}{2r} \approx (1-r^{p'})\frac{\mathbb{E}[B\mathbf{1}_{B>a^{p'}}]}{a^{p'} - r^{p'} p'} + F_\psi(a^{p'}).$$

We shall have thus, $r^{p'} p' < a^{p'} < p'$.

Under Assumption 42, the average conditional expected gain $G^{p'}(x)$ when $x < p'$ can be approximated by:

$$G^{p'}(x) \approx x - r^{p'} p' - \frac{r(1-r^{p'})\mathbb{E}[B\mathbf{1}_{B>x}]}{1 - rF_\psi(x) - (1-r)F_{\kappa^u}(L(x))}. \tag{IV.14}$$

Replacing Equation IV.14 in the deduction of Theorem 6, we have:

Theorem 7. *If the informed trader is assumed to react to the new profitable liquidities with some exponentially distributed reaction time (with intensity λ^a) and the efficient price $P(t-)$ is assumed to be known to the market makers, then under Assumption 31, 33 and 41, immediately after a buying transaction happens at the price $P(t-)+p'$ at time t, the cumulative LOB shape function satisfies: for every $x \in \mathbb{R}^+$, if $L(x) > 0$ and $l(x) > 0$, then $G^{p'}(x) = 0$. Moreover, if Assumption 42 applies, then the best ask price $a^{p'}$ immediately after the transaction satisfies:*

$$r^{p'} p' < a^{p'} < p'.$$

And the cumulative LOB shape function $L(x)$ can be approximated by the following equation when $x < p'$:

$$L(x) \quad \approx \quad F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r(1-r^{p'})\mathbb{E}[\max(B,x)] - r r^{p'}(p'-x)F_{\psi}(x)}{(1-r)(x-r^{p'}p')}\right).$$

In particular, we have, for $x = p'$:

$$\begin{aligned} L(p') &\approx F_{\kappa^u}^{-1}\left(\frac{1}{1-r} - \frac{r}{1-r}\mathbb{E}[\max(\frac{B}{p'},1)]\right) \\ &= L^-(p'), \end{aligned}$$

with $L^-(p')$ the cumulative LOB shape function right before the transaction at the price $P(t-)+p'$.

Theorem 7 suggests that in such settings, the cumulative number of orders immediately after a trade event at p' stayes almost the same, while there will be no quantities available between $[p', p'+\delta p)$ for some $\delta p > 0$, as the condition $G^{p'}(x) = 0$ cannot be satisfied for $x \in [p', p'+\delta p)$ due to the discontinuity of the value function at p'. The question about how the LOB recovers to its previous state or converges to a new equilibrium state will be discussed latter.

Power-Law Distributed Information

We give here an example of power-law distributed information. When $f_{\psi}(x) = Cx^{-\gamma}$ for $x \in (\epsilon, \infty)$, we can compute:

$$\begin{aligned} \mathbb{E}[B|B > p'] &= \int_{p'}^{\infty} \frac{x f_{\psi}(x)}{1-F_{\psi}(p')} dx \\ &= \int_{p'}^{\infty} \frac{x^{-\gamma+1}(\gamma-1)}{p'^{-\gamma+1}} dx \\ &= \frac{(\gamma-1)p'}{(\gamma-2)}. \end{aligned}$$

Thus:

$$r^{p'} = \frac{\gamma-2}{\gamma-1}.$$

And:

$$\begin{aligned}
\frac{1+r}{2r} &\approx \frac{C(a^{p'})^{2-\gamma}}{\frac{\gamma-1}{\gamma-2}a^{p'}-p'}+1-\frac{C(a^{p'})^{1-\gamma}}{\gamma-1}\\
\frac{1-r}{2r} &\approx \frac{C(a^{p'})^{2-\gamma}}{\frac{\gamma-1}{\gamma-2}a^{p'}-p'}-\frac{C(a^{p'})^{1-\gamma}}{\gamma-1}\\
\frac{1-r}{2r} &\approx C(a^{p'})^{1-\gamma}(\frac{a^{p'}}{\frac{\gamma-1}{\gamma-2}a^{p'}-p'}-\frac{1}{\gamma-1}).
\end{aligned}$$

Let $a^{p'}=r^{p'}p'+x$, then the best ask price $a^{p'}$ satisfies the following equation in this example:

$$\frac{1-r}{2r} \approx C(a^{p'})^{1-\gamma}(\frac{r^{p'}p'}{x}+r^{p'}).$$

LOB Recovery Speed

Consider the average expected conditional gain of passive selling orders placed at x at time $t+\delta t$ (denoted by $G^{p'}(x,\delta t)$), given the fact that no trade arrives between $(t,t+\delta t]$. We have, similar to Equation IV.13, for $x<p'$:

$$G^{p'}(x,\delta t)=r^{p'}_{\delta t}\tilde{G}^{p'}_{\delta t}(x)+(1-r^{p'}_{\delta t})G_{\delta t}(x),$$

with $r^{p'}_{\delta t}$ representing the conditional probability that the last transaction is sent by the informed trader, given the fact that no trade arrives between $(t,t+\delta t)$:

$$\begin{aligned}
r^{p'}_{\delta t} &= r^{p'}\frac{e^{-(\lambda^i+\lambda^u+\lambda^a)\delta t}}{r^{p'}e^{-(\lambda^i+\lambda^u+\lambda^a)\delta t}+(1-r^{p'})e^{-(\lambda^i+\lambda^u)\delta t}}\\
&= \frac{1}{1+\frac{1-r^{p'}}{r^{p'}}e^{\lambda^a\delta t}},
\end{aligned}$$

and $\tilde{G}^{p'}_{\delta t}(x)$ the average expected conditional gain of passive selling orders placed at x at time $t+\delta t$ in the case that the last transaction is an informed trade, $G_{\delta t}(x)$ that in the case of a noise trade. It is easy to see that $G_{\delta t}(x)=G(x)$ and $\tilde{G}^{p'}_{\delta t}(x)=\tilde{G}^{p'}(x)$, thus, for $x<p'$:

$$G^{p'}(x,\delta t)=r^{p'}_{\delta t}\tilde{G}^{p'}(x)+(1-r^{p'}_{\delta t})G(x).$$

When $\delta t\to\infty$, $r^{p'}_{\delta t}\to 0$, and $G^{p'}(x,\delta t)\to G(x)$, that is, if the trade is intiated by the noise trader, the LOB shape will gradually recover after this trade to its stationary state, with speed $\sim e^{-\lambda^a t}$.

6 Conclusion and perspectives

In this chapter, we introduce an agent-based model on the LOB dynamics. Based on a greedy assumption on the behavior of the informed trader and a zero-profit assumption on the market makers, a link between the proportion of noise trade, bid-ask spread, dynamics of the efficient price and the equilibrium LOB state is established. The effect of introducing the tick value constraint is then discussed, and we show that the constrained bid-ask spread is equal to the sum of the tick value and the intrinsic bid-ask spread that corresponds to the case of zero tick value. Price discretization also enables the explicit computation of the priority value when the tick value is positive. The model is then extended to study the consequence of adding the uncertainty on the efficient price on the market makers. Two main subjects are discussed in this generalized model: the information propagation between the informed trader and the market makers and the LOB recovery speed.

Although many interesting discoveries have been made in this chapter, a lots remain to be done. Till now, only market makers are allowed to insert limit orders, this is clearly unrealistic. In real market, the roles of the informed trader and that of the market makers are often mixed, as the informed trade has also the possibility of placing passive limit orders. By placing limit orders to the market, the informed trader may leak some information to the market makers, which may lead to an adjustment of their estimated value of the efficient price. The consequences of introducing such interactions are still unknown to us.

Bibliography

Abergel, F. and Jedidi, A. (2011), "A mathematical approach to order book modelling," in *Econophysics of Order-driven Markets*, Springer, pp. 93–107.

— (2015), "Long Time Behaviour of a Hawkes Process-Based Limit Order Book," *Available at SSRN 2575498.*

Ahn, H.-J., Cao, C. Q., and Choe, H. (1996), "Tick size, spread, and volume," *Journal of Financial Intermediation*, 5, 2–22.

Almgren, R. and Chriss, N. (2000), "Optimal execution of portfolio transactions," *Journal of Risk*, 3, 5–39.

Avellaneda, M. and Stoikov, S. (2008), "High-frequency trading in a limit order book," *Quantitative Finance*, 8, 217–224.

Bacidore, J. M. (1997), "The impact of decimalization on market quality: An empirical investigation of the Toronto Stock Exchange," *Journal of Financial Intermediation*, 6, 92–120.

Bayer, C., Horst, U., and Qiu, J. (2014), "A Functional Limit Theorem for Limit Order Books with State Dependent Price Dynamics," *arXiv preprint arXiv:1405.5230.*

Bean, N. and Latouche, G. (2010), "Approximations to quasi-birth-and-death processes with infinite blocks," *Advances in Applied Probability*, 42, 1102–1125.

Bertsimas, D., Lo, A. W., and Hummel, P. (1999), "Optimal control of execution costs for portfolios," *Computing in Science and Engg.*, 1, 40–53.

Bessembinder, H. (2000), "Tick size, spreads, and liquidity: An analysis of Nasdaq securities trading near ten dollars," *Journal of Financial Intermediation*, 9, 213–239.

Billingsley, P. (2009), *Convergence of probability measures*, vol. 493, John Wiley & Sons.

Bini, D., Meini, B., Steffé, S., and Van Houdt, B. (2006), "Structured Markov chains solver: algorithms," in *Proceeding from the 2006 workshop on Tools for solving structured Markov chains*, ACM, p. 13.

Bouchard, B., Dang, N.-M., and Lehalle, C.-A. (2011), "Optimal control of trading algorithms: a general impulse control approach," *SIAM Journal on Financial Mathematics*, 2, 404–438.

Bouchaud, J.-P., Mézard, M., Potters, M., et al. (2002), "Statistical properties of stock order books: empirical results and models," *Quantitative finance*, 2, 251–256.

Bourghelle, D. and Declerck, F. (2004), "Why markets should not necessarily reduce the tick size," *Journal of banking & finance*, 28, 373–398.

Chung, K. H. and Chuwonganant, C. (2002), “Tick size and quote revisions on the NYSE,” *Journal of Financial Markets*, 5, 391–410.

Chung, K. H. and Van Ness, R. A. (2001), “Order handling rules, tick size, and the intraday pattern of bid–ask spreads for Nasdaq stocks,” *Journal of Financial Markets*, 4, 143–161.

Cont, R. and De Larrard, A. (2013), “Price dynamics in a Markovian limit order market,” *SIAM Journal on Financial Mathematics*, 4, 1–25.

Cont, R., Stoikov, S., and Talreja, R. (2010), “A stochastic model for order book dynamics,” *Operations research*, 58, 549–563.

Dayri, K. and Rosenbaum, M. (2012), “Large tick assets: implicit spread and optimal tick size,” *Preprint arXiv:1207.6325*.

Delattre, S., Robert, C. Y., and Rosenbaum, M. (2013), “Estimating the efficient price from the order flow: a Brownian Cox process approach,” *Stochastic Processes and their Applications*, 123, 2603–2619.

Farmer, J. D., Gerig, A., Lillo, F., and Waelbroeck, H. (2013), “How efficiency shapes market impact,” *Quantitative Finance*, 13, 1743–1758.

Foucault, T. (1999), “Order flow composition and trading costs in a dynamic limit order market,” *Journal of Financial markets*, 2, 99–134.

Gareche, A., Disdier, G., Kockelkoren, J., and Bouchaud, J.-P. (2013), “A Fokker-Planck description for the queue dynamics of large tick stocks,” *Preprint arXiv:1304.6819*.

Gatheral, J. (2010), “No-dynamic-arbitrage and market impact,” *Quantitative Finance*, 10, 749–759.

Glosten, L. R. and Milgrom, P. R. (1985), “Bid, ask and transaction prices in a specialist market with heterogeneously informed traders,” *Journal of financial economics*, 14, 71–100.

Goldstein, M. A. and Kavajecz, K. A. (2000), “Eighths, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE,” *Journal of Financial Economics*, 56, 125–149.

Gross, D. and Harris, C. M. (1998), *Fundamentals of queuing theory*, Wiley New York.

Harris, L. and Hasbrouck, J. (1996), “Market vs. limit orders: the SuperDOT evidence on order submission strategy,” *Journal of Financial and Quantitative analysis*, 31, 213–231.

Ho, T. and Stoll, H. R. (1981), “Optimal dealer pricing under transactions and return uncertainty,” *Journal of Financial economics*, 9, 47–73.

Huang, R. D. and Stoll, H. R. (1997), “The components of the bid-ask spread: A general approach,” *Review of Financial Studies*, 10, 995–1034.

Huang, W., Lehalle, C.-A., and Rosenbaum, M. (2013), “Simulating and analyzing order book data: The queue-reactive model,” *arXiv preprint arXiv:1312.0563*.

Jacod, J. and Shiryaev, A. (2013), *Limit theorems for stochastic processes*, vol. 288, Springer Science & Business Media.

Lachapelle, A., Lasry, J.-M., Lehalle, C.-A., and Lions, P.-L. (2013), "Efficiency of the price formation process in presence of high frequency participants: a mean field game analysis," *arXiv preprint arXiv:1305.6323.*

Lakner, P., Reed, J., and Stoikov, S. (2013), "High frequency asymptotics for the limit order book," *Preprint.*

Laruelle, S., Lehalle, C.-A., and Pagès, G. (2013), "Optimal posting price of limit orders: learning by trading," *Mathematics and Financial Economics*, 7, 359–403.

Latouche, G. and Ramaswami, V. (1999), *Introduction to matrix analytic methods in stochastic modeling*, vol. 5, ASA-SIAM.

Lau, S. T. and McInish, T. H. (1995), "Reducing tick size on the Stock Exchange of Singapore," *Pacific-Basin Finance Journal*, 3, 485–496.

Lehalle, C.-A., Laruelle, S., Burgot, R., Pelin, S., and Lasnier, M. (2014), *Market Microstructure in Practice*, World Scientific publishing.

Madhavan, A., Richardson, M., and Roomans, M. (1997), "Why do security prices change? A transaction-level analysis of NYSE stocks," *Review of Financial Studies*, 10, 1035–1064.

Meyn, S. P. and Tweedie, R. L. (1993), "Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes," *Advances in Probability*, 518–548.

— (2009), *Markov chains and stochastic stability*, Cambridge University Press.

Moallemi, C. and Yuan, K. (2015), "The value of queue position in a limit order book," *Working paper.*

Norris, J. R. (1998), *Markov chains*, no. 2008, Cambridge University Press.

Parlour, C. A. (1998), "Price dynamics in limit order markets," *Review of Financial Studies*, 11, 789–816.

Robert, C. Y. and Rosenbaum, M. (2011), "A new approach for the dynamics of ultra-high-frequency data: the model with uncertainty zones," *Journal of Financial Econometrics*, 9, 344–366.

— (2012), "Volatility and covariation estimation when microstructure noise and trading times are endogenous," *Mathematical Finance*, 22, 133–164.

Roşu, I. (2009), "A dynamic model of the limit order book," *Review of Financial Studies*, 22, 4601–4641.

Smith, E., Farmer, D. J., Gillemot, L., and Krishnamurthy, S. (2003), "Statistical theory of the continuous double auction," *Quantitative Finance*, 3, 481–514.

Toth, B., Lemperiere, Y., Deremble, C., De Lataillade, J., Kockelkoren, J., and Bouchaud, J.-P. (2011a), "Anomalous price impact and the critical nature of liquidity in financial markets," *Physical Review X*, 1, 021006.

Toth, B., Palit, I., Lillo, F., and Farmer, J. D. (2011b), "Why is order flow so persistent?" *arXiv preprint arXiv:1108.1632.*

Wu, Y., Krehbiel, T., and Brorsen, B. W. (2011), "Impacts of tick size reduction on transaction costs," *International Journal of Economics and Finance*, 3, p57.

Wyart, M., Bouchaud, J.-P., Kockelkoren, J., Potters, M., and Vettorazzo, M. (2008), "Relation between bid–ask spread, impact and volatility in order-driven markets," *Quantitative Finance*, 8, 41–57.